





Délégation aux Systèmes d'Information
Pôle des "services d'appui à la recherche"
**Informatique scientifique**

## Appel à projets interne

# SPIRALES 2010

## Formulaire de demande DSI-SPIRALES

## « Soutien aux Projets Informatiques dans les Equipes Scientifiques »

3ème cas: Nouveau projet en calcul intensif et modélisation (projet finalisé de mise en place d'un calculateur ou développement d'une application IS)

Remise des projets :

21 décembre 2009

à spirales@ird.fr

*Contact DSI :*

**Equipe Informatique Scientifique**

equipe-is@ird.fr

# Demande d'un soutien DSI sur les projets informatiques des UR/US.

## Formulaire de demande DSI-SPIRALES 2010
## « Soutien aux Projets Informatiques dans les Equipes Scientifiques ».

**Le présent formulaire** comporte différentes parties, avec des questions parfois très précises. Il est destiné à **faciliter le travail des évaluateurs**.

Les propositions doivent être adressée sous forme électronique (au format RTF, DOC ou PDF) à l'adresse suivante : spirales@ird.fr

**4 cas sont possibles dans le cas d'un projet SPIRALES 2010.**

Projet SPIRALES 2010

4 cas possibles :

Etude de faisabilité « calcul intensif/modélisation » (définition des besoins)

Nouveau projet (projet finalisé)

Continuum d'un projet existant

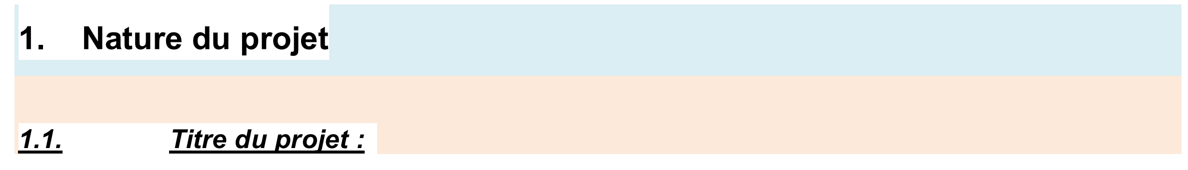
Développement d'une application IS « BD/SIG/Application web/Algo … » (Suite à une étude de faisabilité financée en 2009)

Développement d'une application IS « Calcul intensif/Modélisation »

| | Parties du formulaire à renseigner selon le cas | 1er cas | 2nd cas | 3ème cas | 4ème cas |
|---|---|---|---|---|---|
| 1. | Nature du projet | Oui | Oui | Oui | Oui |
| 2. | Porteur(s) de projet | Oui | Oui | Oui | Oui |
| 3. | Moyens / appui demandés à la DSI | Oui | Oui | Oui | Oui |
| 4. | Bilan / Etat d'avancement des phases précédentes | Non | Non | Non | Oui |
| 5. | Description des besoins | Oui | Oui | Oui | Non |
| 6. | Description du projet de développement d'application IS (méthodes, solutions, et moyens) | Non | Oui | Oui | Oui (*) |
| 7. | Description du projet hors développement (méthodes, solutions, et moyens) | Non | Non | Non | Non |
| 8. | Pertinence, résultats/livrables attendus et valorisation du projet | Oui | Oui | Oui | Oui |

(*) Dans le cas d'un **continuum**, si le formulaire 2009 a déjà été détaillé, la **partie 6 peut être remplacée par tous documents utiles**: dossier d'expression des besoins ou cahier des charges, devis détaillé…

**Ce formulaire est spécifique au <u>3ème cas: Nouveau projet en calcul intensif et modélisation (projet finalisé de développement d'application IS).</u>**



# 1. Nature du projet

## ***1.1. Titre du projet :***

EPIS : Easy, Parallel and Interactive Sworm

## ***1.2. Résumé du projet proposé (5 lignes maximum)***

Le projet présentement défendu vise à permettre la réalisation de simulations Sworm via le web. Il s'agit de mettre en place un portail web, didactique couplé à une infrastructure de calcul haute performance en vue de permettre la réalisation d'expérimentations et de plans d'expériences autour du modèle Sworm avec des temps de simulation acceptables. Ce projet tend vers la construction d'une plate-forme de simulation pluridisciplinaire permettant et

facilitant l'accès aux outils de calcul haute performance grâce à l'utilisation des technologies du web et des systèmes distribués.

### 1.3. Type de projet

| | |
|---|---|
| 1er cas | þ Etude de faisabilité (définition des besoins) pour un projet en calcul intensif / modélisation : Demande d'appui pour une analyse fine des besoins et la formulation de spécifications, éventuellement développement d'un prototype (en vue d'une seconde phase destinée au développement et à la réalisation du projet) |
| 2nd cas | o **Projet finalisé de développement d'une application IS** (proposition finalisée et détaillée en matière d'expression des besoins, d'identification des solutions et des moyens…) **suite à une étude de faisabilité financée en 2009.** |
| 3ème cas | **Nouveau projet SPIRALES :** ☑ **Projet finalisé de mise en place d'un calculateur et de développement d'une application IS en calcul intensif et modélisation** (proposition finalisée et détaillée en matière d'expression des besoins, d'identification des solutions et des moyens…) |
| 4ème cas | o **Continuum d'un projet SPIRALES existant** (prévu sur 2008-2009-2010 ou suite d'un précédent projet SPIRALES)<br>Continuum d'un **développement d'une application IS**<br>o continuum d'un projet **autre qu'un développement** d'application IS |

Les demandes d'hébergement d'applications IS, d'accès à un serveur de développement, de création de dépôt Subversion (SVN), de formations IS… ne constituent pas des demandes SPIRALES et doivent être adressées directement à equipe-is@ird.fr sans échéance particulière.

### 1.4. Durée prévue :

☐ Durée prévue : ☐ 1 an ✠ 2 ans

## 2. Porteur(s) de projet

### 2.1. Unité :

x UMI ☐ UR ☐ US N° 209 Nom : UMMISCO

### ***2.2. Département***

x DME ☐ DRV ☐ DSS

### ***2.3. Nom du porteur de projet :***

Marilleau Nicolas

### ***2.4. Statut et coordonnées du porteur de projet :***

Titulaire IR

32 avenue Henri Varagnat

93143 Bondy CEDEX

tel 0148027901/ Fax : 01 48 02 55 43 / nicolas.marilleau@ird.fr

### ***2.5. Nom et coordonnées du Directeur d'Unité (si différent) :***

Pierre Auger

Titulaire DR

Maroc

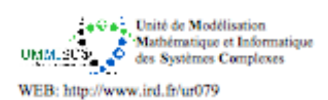
Tel : 01 48 02 55 24 / Fax : 01 48 02 55 43 / pierre.auger@bondy.ird.fr

### ***2.6. Aval du directeur d'unité (obligatoire).***

*OU adressé par mail à spirales@ird.fr*




Centre de recherche d'Ile-de France
32, avenue Henri Varagnat
F-93143 Bondy cedex. France.
TEL : 33(0)1 48 02 55 00
FAX : 33(0)1 48 47 30 88
WEB : www.bondy.ird.fr



WEB: http://www.ird.fr/ur079

**Professeur Pierre AUGER**
Membre de l'Académie des Sciences
Françaises
Directeur de l'UMI UMMISCO 209
TEL direct : 33(0)1 48 02 59 94
FAX : 33(0)1 48 47 30 88
E-MAIL : auger@bondy.ird.fr

Le projet Spirales EPIS a pour objet de créer un environnement web qui donne la possibilité à des thématiciens de réaliser des simulations distribuées hautes performances. Ce projet permettra, d'une part, de valoriser quelques modèles développés par les chercheurs de Ummisco, de l'IRD et d'autres instituts, et d'autre part d'initier un groupe de thématiciens aux techniques de modélisation-simulation.

Le projet EPIS est de première importance pour l'UMI UMMISCO et, au-delà, pour promouvoir les techniques de modélisation mathématique et informatique au sein de l'IRD. De plus, il s'inscrit parfaitement dans la démarche d'UMI UMMISCO et permettra de renforcer la collaboration avec les partenaires du Sud.

Aux vues des objectifs du projet et des résultats déjà obtenus dans le domaine, j'apporte donc mon soutien à ce projet porté par les chercheurs informaticiens de notre laboratoire.

Fait à Bondy, le 23 décembre 2009

P. AUGER

Directeur de l'UMI UMMISCO 209
Membre de l'Académie des Sciences française

### ***2.7. Implantation principale de l'unité :***

UMMISCO, Bondy

### ***2.8. Site(s) de déroulement du projet :***

- **UMMISCO (dir. Pierre Auger) IRD- Bondy**
- **ENSAI (Pr. Carl MBOFUNG ) – Université de Ngaoundéré, Cameroun**
- **IFI/MSI (dir. Ho Tuong Vinh) – Hanoi, Vietnam**
- **LIFC (dir. Jean Christophe Lapayre) Université de Franche Comté, Besançon**
- **LIRMM (dir. Michel Robert) Université de Montpellier, Montpellier**

### ***2.9. Site administratif à partir duquel se feront les dépenses budgétaires***

1800 € à la régie du Vietnam, le reste à Bondy

(*attention: cette information doit être bien renseignée car elle évite les disfonctionnements observés quand il s'agit de transférer les crédits d'un site à un autre en cours d'année*)

### ***2.10. Projets inter-unité ou inter-organismes :***

**☐ Projet inter-unités x Projet inter-organismes**

### ***2.11. Liste des unités ou organismes partenaires du projet***

- **IRD - UMMISCO - dir. Pierre Auger – Bondy**
- **Université de Franche Comté - LIFC - Jean Christophe Lapayre – Besançon**
- **Université de Montpellier - LIRMM - Michel Robert – Montpellie**
- **Université de Ngaoundéré, Camerour**
- **IFI - MSI - Ho Tuong Vinh – Hanoi, Vietnam**

### ***2.12. Liste des intervenants impliqués de manière effective dans la réalisation du projet :***

Eric Blanchart - DR IRD - Eco&Sol - Expertise pour la conception d'une IHM adaptée aux sciences du sol

Christophe Cambier - MCF- UMMISCO - Distribution du simulateur Sworm

Clive Canape - Titulaire IR- DSI- Montpellier Suivit de prestation et conception

Jacques Ferber - Pr - LIRMM - Montpellier - Madkit

Benoit Gaudou - post-doc - Hanoi - Mise en place de l'interface graphique

Jean Claude Kamgang Equivalent MCF - Cameroun - Expertise pour la conception d'une IHM

Tuong Vinh Ho- Equivalent MCF - Hanoi - Mise en place de l'interface graphique

Christophe Lang - MCF - LIFC - Besançon - Rafale-sp, Algorithme distribué

Nicolas Marilleau - Titulaire IR - UMMISCO - Bondy - Algorithme distribué, distribution de Sworm, Porteur du projet

Fabien Michel - MCF - LIRMM - Montpellier - Madkit
Laurent Philippe - Pr - LIFC - Besançon - Exécution cluster

### ***2.13. Disponibilité / implication de chacun des intervenants effectifs : exprimée en % de temps-homme ou en jours-homme (ETP total ou pour une période)***

Eric Blanchart 10% ETP sur la durée du projet

Christophe Cambier 10% ETP sur la durée du projet

Clive Canape 10% ETP sur la durée du projet

Jacques Ferber 10% ETP sur la durée du projet

Benoit Gaudou 15% ETP sur la durée du projet

Tuong Vinh Ho 15% ETP sur la durée du projet

Jean Claude Kamgang 10% ETP sur la durée du projet

Christophe Lang 10% ETP sur la durée du projet

Nicolas Marilleau 25% ETP sur la durée du projet

Fabien Michel 10% ETP sur la durée du projet

Laurent Philippe 10% ETP sur la durée du projet

## 3. Moyens / appui demandés à la DSI

### ***3.1 Contribution financière demandée à la DSI pour 2010 en euros HT :***

**Montant 2010 demandé : 15000 € HT soit**

Ventilation par poste :

Fonctionnement : 4000

Equipement : 2000

Prestation de service : 9000

*(les informations apportées doivent être cohérentes avec celles précisées à la question 24.)*

### ***3.2 Demande envisagée pour 2011 – si projet de 2 ans - en euros HT :***

**Montant 2011 envisagé : 15500 € HT soit**

Ventilation par poste :

Fonctionnement : 6500

Equipement : 0

Prestation de service : 9 000

### ***3.3 Montant(s) précédemment attribué(s) par la DSI - en euros HT :***

| | **2005** | **2006** | **2007** | **2008** | **2009** |
|---|---|---|---|---|---|
| Montants attribués (€ HT) | 0 | 0 | 0 | **0** | **0** |

### 3.4 Moyens affectés au projet et Cofinancements acquis hors SPIRALES (€ HT) :

Autres sources de financements acquis : Montant (€ HT) :

Moyens apportés par l'unité (hors ressources humaines) Montant (€ HT) : 488 000 €

Les moyens apportés sont :

- Le cluster 360 000 Euros
- Les logiciels pour le cluster 68 000 Euros
- Le coût de fonctionnement (électricité, maintenance, ...) 30 000 Euros par an

### 3.5 Moyens humains affectés au projet :

Total des moyens humains affectés au projet par les unités et partenaires (exprimé en total de jours-homme ou ETP (Equivalent Temps Plein)) (*cf. définition et exemple à la question 17.*) :

les moyen humain s'élève à 135 ETP. Ces moyen réaliserons principalement la conception et l'encadrement des équipes de développement. Le développement sera réalisé dans le cadre de prestation lors de stage.

### 3.6 Coût total estimé du projet (toutes années confondues) :

Estimation du coût total du projet toutes années SPIRALES confondues : crédits SPIRALES, moyens fournis par l'unité et cofinancements acquis (hors ressources humaines) : 503500 € HT

### 3.7 Ressources humaines extérieures mobilisées ou demandées :

☐ Compétences mobilisées ou souhaitées (profil type) :

✠ Intervention d'un/de prestataire(s) de service :

Nguyen Tuan Thanh Le, étudiant en master qui a déjà fait ses preuves dans le cadre d'autres projets de recherche

Mathieu Gandon, Etudiant en master 2 à l'université de Montpellier

X Mobilisation d'un/de stagiaire(s) *(sous réserve de compétences fortes en informatique scientifique au sein de l'équipe porteur du projet et de capacités de l'équipe à dégager du temps pour assurer un réel encadrement)*

✠ Demande d'appui de l'équipe 'Informatique scientifique' de la DSI / pour l'appui méthodologique et le suivi de projet : rédaction de cahiers des charges, suivi des prestations (respect des plannings, vérification des livrables...) et assistance à l'encadrement de stagiaires

✠ Demande d'appui de l'équipe 'Informatique scientifique' de la DSI / pour le développement et/ou la réalisation du projet (avec estimation du temps-homme nécessaire) :

assistance à "l'étude de l'architecture" de l'interface entre PAMS et un cluster : 10 ETP ;

> *La DSI, suite au comité d'évaluation, pourra pour quelques projets et sur quelques sites (Nouméa, Dakar, Montpellier…) et dans la limite des* ***moyens humains de la DSI disponibles****, convertir ces demandes d'appui ou de financement de prestataire de service en temps-homme, c'est-à-dire par une intervention directe du 'pool informatique scientifique'.*

### 3.8 Demande d'un dépôt Subversion (SVN) :

*Description des besoins pour ce projet SPIRALES (une demande formelle et détaillée, avec signature de la charte sera néanmoins nécessaire dans un 2nd temps) - (Définition SVN:* http://fr.wikipedia.org/wiki/Subversion_(logiciel)*)*

Nous ne demanderons pas de zone de dépôt svn spécifique au projet. Cependant, nous demanderons la création d'un sous projet svn appartenant à https://svn.mpl.ird.fr/ur079geodes. La gestion de ce sous projet sera confié au porteur de ur079geodes (Nicolas Marilleau)

### 3.9 Demande d'hébergement(s) / d'accès à un (des) serveur(s) 1/ de développement et de tests pour la durée du projet, 2/ de 'pré production' et de recette pendant ou à l'issue du projet, 3/ d'exploitation à l'issue du projet :

*Description des besoins pour ce projet SPIRALES : technologies, capacité… (une demande formelle et détaillée, avec signature de la charte sera néanmoins nécessaire dans un 2nd temps)*

Les outils (un cluster) qui seront utilisés au sein de ce projet n'existe pas dans l'offre actuelle de la DSI. Nous nous reposerons donc sur les ressources de calcul proposées par l'Université de Franche-Comté.

### 3.10 Appui de la DSI apporté pour l'élaboration du projet ?

*Si vous avez bénéficié de l'appui de la DSI (coordination IS, pool d'informaticiens scientifiques de Dakar ou Nouméa, SIL…) pour l'élaboration de cette proposition, décrivez très brièvement le type d'appui.*

Une réunion explicative sur "l'appel à projets en calcul intensif et modélisation", ainsi qu'une réunion concernant les grands axes du projet, ont eu lieu respectivement en octobre et novembre avec un membre de l'Informatique Scientifique.

## 5. Description des besoins

La demande peut-être être **accompagnée de tous documents utiles** :

*présentation du projet global ou descriptif du projet, rapport de phases préliminaires, étude de faisabilité (définition des besoins), dossier d'expression des besoins ou **cahier des charges**, devis détaillé…*

### 5.1 Objectifs scientifique

#### *5.1.a Préambule*

Les sols jouent un rôle majeur dans la fourniture des services et biens fournis aux humains par les écosystèmes. Situés à l'interface entre l'atmosphère, la lithosphère, l'hydrosphère et la biosphère, ils participent aux grands cycles nécessaires à la vie sur Terre : cycle de l'eau et des nutriments majeurs (carbone, azote, phosphore…). Ils supportent la plupart des systèmes de production agricoles, sylvicoles et pastoraux et participent à la régulation du climat (en contrôlant les émissions de gaz à effet de serre et la séquestration du carbone), au contrôle de l'érosion et à la détoxification. L'étude des sols contribue donc à une préservation des écosystèmes et à l'amélioration des politiques pastorales dans les pays du Sud, ce qui constitue un des enjeux majeurs de la recherche actuelle menée par les universités et les instituts comme l'IRD.

Les sols sont reconnus comme étant des systèmes biologiques et écologiques particulièrement complexes : ils sont en effet composés d'un grand nombre d'entités qui entretiennent de multiples interactions entre elles et avec leur environnement. Pour comprendre et prévoir leur fonctionnement, des besoins en approches systémiques et de modélisation sont de plus en plus prégnants. Pour cela, des moyens informatiques (conceptuels et logiciels) spécifiques au domaine sont nécessaires. Pourtant, ils demeurent pratiquement inexistants en science du sol.

Des modèles destinés à décrire et comprendre le fonctionnement du sol et les organismes du sol existent. Ils représentent généralement les écosystèmes sols de manière globale et agrégée pour contourner la complexité liée à la description des nombreuses interactions existant entre les organismes et entre les organismes et leur environnement,

et à la représentation de l'hétérogénéité spatiale du sol (Smith et al., 1998). Pourtant, dans certains cas, la prise en compte explicite des groupes fonctionnels d'organismes permettrait de mieux décrire le fonctionnement du sol (Groffman & Jones, 2000). Un modèle multi-agents (nommé SWORM), conçu au sein de l'IRD par une équipe de biologistes (UMR Eco&Sol) et de modélisateurs (UMI UMMISCO), existe depuis peu pour décrire l'effet des vers de terre sur la structure du sol (Marilleau et al., 2008 ; Blanchart et al., 2009). Ce modèle permet entre autres de suivre l'organisation spatiale des déjections de vers de terre dans le sol ainsi que la dynamique de la matière organique. Les résultats obtenus par ce nouveau modèle sont novateurs et présentent un intérêt important pour la communauté de sciences du sol. L'article (Blanchart et al., 2009) était en 2009 l'un des 10 articles les plus lus de la revue *European Journal in Soil Sciences,* revue qui fait référence en sciences du sol. Mais la mise en place d'un tel simulateur et la réalisation des simulations ont nécessité l'emploi de ressources informatiques considérables : **la conception et l'installation du simulateur nécessite le concours d'un information modélisateur ; une simulation Sworm nécessite environ une semaine** de calcul avant d'observer une stabilisation (résultats obtenus sur un poste de travail avec des performances moyennes). Dans de telles conditions, l'impact pour la **communauté** de Sworm est **bridé par sa complexité d'installation et d'utilisation**. Il n'est pas possible pour un chercheur lambda d'imaginer : (i) télécharger Sworm à partir d'un site web ; (ii) installer le simulateur sur son poste de travail ; et (iii) réaliser des expérimentations d'une semaine qu'il faudra refaire car, ne connaissant pas le simulateur, le chercheur ne pourra de toute façon pas le calibrer dès la première utilisation.

Au-delà du modèle Sworm, **les besoins scientifiques actuels en termes de représentation de la complexité sont de plus en plus importants**. La conception de simulateurs « jouets », ad hoc et facilement échangeables cède la place à la création de simulateurs plus importants souvent connectés à des bases de données (parfois géographiques) et nécessitant une puissance de calcul considérable pour leurs exécutions. Un des enjeux actuels du monde scientifique et des instances comme la DSI se situe dans **la promotion d'outils qui facilitent l'accès à la simulation à haute performance** pour des candides de l'informatique, en se reposant sur les technologies du web, de la simulation et des systèmes répartis et en utilisant des infrastructures informatiques à forte puissance tels que des clusters ou des grilles.

***5.1.b Intérêt de la simulation haute performance orientée web***

Depuis de nombreuses années le recours au calcul haute performance permet l'exploration de scénarios avec des ordres de grandeur bien supérieurs à ce qui peut être fait sur un ordinateur classique. Dans le cas d'une étude paramétrique, deux approches peuvent être mises en oeuvre pour permettre l'utilisation de ressources plus puissantes : (i) la parallélisation de l'expérimentation ; (ii) la distribution des simulateurs.

La parallélisation de l'expérimentation consiste à exécuter, sur des noeuds différents, des instances de simulation avec des paramètres différents. Cette approche ne nécessite aucune modification du simulateur, si ce n'est son déployement dans le contexte particulier du cluster. Par contre, elle suppose, avant de lancer une campagne de tests, d'avoir une bonne idée des paramètres pertinants. Il est alors parfois nécessaire d'attendre une semaine, le temps de terminer le jeu de tests courants, pour préparer un second jeu de tests. Cela suppose aussi d'avoir suffisamment de paramètres dans les jeux de test, permettant ainsi plusieures exécutions concurrentes, pour profiter du grand nomlbre de ressources. Dans ce cas, une démarche incrémentale visant à rafiner successivement les paramètres, expérimentation après expérimentation, n'est pas envisageable.

La distribution des simulateurs consiste à modifier le simulateur de manière à ce qu'il exploite les ressources du cluster en une seule exécution. Ici aussi deux méthodes sont applicables : une dite "en mémoire partagée" et la seconde en "mémoire distribuée". Dans la première, le simulateur exploite le parallèlisme des nouvelles architectures de processeurs, les coeurs, et les architecture multi-processeurs. Dans la seconde, le simulateur exploite un ensemble de noeuds, en divisant son espace de travail pour coller à l'architecture de la machine. Il faut noter que l'architecture distribuée se prête normalement bien à la parallèlisation d'application travaillant sur un domaine décomposable : il faut alors répartir le domaine entre différentes machines et gérer la cohérence des frontières.

L'avènement des technologies du web permet d'envisager de nouvelles méthodes d'accès et d'utilisation des ressources de calcul à haute puissance pour la réalisation de simulations. Grâce à elle, il est possible d'imaginer des passerelles humanisées, orientés thématiciens se concrétisant sous la forme de simples interfaces web mais qui permettent l'accès à des ressources (logicielles comme des simulateurs distribués et matérielles telles qu'un cluster)

complexes à manipuler, à configurer, à partager, à capitaliser et à maintenir en fonctionnement. Ce projet s'inscrit justement dans la promotion de ce type d'environnement

***5.1.c Objectif du projet***

Le projet présentement défendu fait suite à une demande d'une équipe de biologistes en sciences du sol (UMR Eco&Sol) qui souhaiterait **rendre plus accessible le modèle Sworm en le publiant sur le web et en permettant la réalisation de simulations**. Sworm fait l'objet d'un intérêt important de la part de la communauté en sciences du sol. Sa mise en ligne améliorerait sa communication et la publicité des techniques de modélisation-simulation, mais surtout, répondrait à de multiples requêtes de chercheurs qui souhaiteraient réutiliser Sworm pour leurs propres recherches : demandes aujourd'hui impossibles à satisfaire compte tenu de la complexité du modèle et des temps de simulation.

Ce projet vise donc à **mettre en place un portail web didactique**, couplé à une infrastructure de **calcul haute performance** en vue de permettre la **réalisation de simulations** du modèle Sworm par des chercheurs en science du sol sans le concours d'informaticiens modélisateurs. Par l'intermédiaire de cet outil, un biologiste sera en mesure de réaliser des expérimentations en temps réel ou de programmer des plans d'expériences qui se réaliseront en différés.

La mise en ligne de Sworm ne constitue pas le point clé de ce projet car des portails comme Pams (Spirales Marilleau & al. 2007- 2009) existent. Le véritable enjeu de ce travail se situe (i) dans la **distribution du simulateur** Sworm et (ii) dans la **mise en place d'algorithmes et d'outils informatiques** qui rendent transparent (du point de vue utilisateur) la réalisation d'expérimentations Sworm sur un cluster.

Les travaux réalisés dans le cadre de ce projet sous le couvert du modèle Sworm, les algorithmes et les outils mis en place ont pour vocation à être généralisés et à pouvoir être appliqués à n'importe quel type de simulateur. Ainsi ce projet tend vers la **construction d'une plate-forme pluridisciplinaire de simulation et d'analyse** qui facilite l'accès à la **simulation haute performance** grâce à l'utilisation des technologies du **web et des systèmes distribués**. L'outil mis en place favorisera la **capitalisation** des connaissances (résultats) et des ressources informatiques (cluster). Les paramètres et les résultats de simulation seront sauvegardés dans une base de données et pourront être ultérieurement réutilisés et commentés par d'autres chercheurs lors de nouvelles expérimentations. Grâce à cet outil les chercheurs auront accès à une formidable puissance de calcul qu'ils ne sont pas en mesure de posséder et d'administrer au sein de leur unité de recherche.

Une dernière phase du projet s'attachera à éprouver l'outil en le soumettant au regard des thématiciens futurs utilisateurs de l'environnement. Ce travail sera réalisé par l'intermédiaire d'une **formation** réalisées en présence de **thématiciens du Nord et du Sud**.

## 5.2 *Description et analyse des besoins*

*OU renvoyer à un document joint*
Comme dit précédement, ce projet vise à déployer le simulateur Sworm sur un cluster et à rendre possible la programmation et la réalisation d'expérimentations, par des spécialistes des sciences du sol, via une interface web orientée thématiciens. Afin de mener à bien ce projet plusieurs points doivent être abordés:

- WP1 Distribution du simulateur Sworm
- WP2 La mise en place d'algorithmes de répartition de charge sur des simulateurs
- WP3 L'optimisation de l'environnement de simulation Madkit pour un déploiement sur cluster
- WP4 La mise en place de routines pour le déploiement et l'exécution de simulateurs sur un cluster de calcul
- WP5 Conception d'une interface web didactique pour la programmation de plans d'expériences et l'exécution de simulateurs sur un cluster
- WP6 formation

Chacun de ces points est individuellement présenté dans la suite de ce document.

**WP1: Distribution du simulateur Sworm (Christophe Cambier, Nicolas Marilleau) –** ***Dakar/Bondy***

Le simulateur Sworm vise à reproduire l'effet des vers de terre sur la structure de l'écosystème sol et donc sur la dynamique de la matière organique. Ce simulateur a été développé en JAVA en se reposant sur la bibliothèque de simulation répartie RAFALES-SP (Marilleau, 2006) qui est elle-même un plugin de l'environnement de simulation MadKit (Ferber et al., J., 2004). Il s'agit d'un simulateur «stand-alone» qui s'exécute sur le poste du chercheur utilisateur. Ce simulateur dispose d'une interface graphique et propose plusieurs sorties dont des graphiques, des vues en 2 ou 3 dimensions, des fichiers de données, des fichiers images et des sorties VRML.

L'objectif de ce WP est : (i) de **construire une nouvelle version du simulateur dépourvue d'interface graphique** ; (ii) **d'améliorer la structure interne du simulateur en vue de le rendre distribuable**.

Dans un premier temps, il s'agit de dissocier l'interface graphique du noyau de simulation afin de construire un nouvel exécutable utilisable sur un serveur dépourvu de serveur graphique. Pour cela, nous nous reposerons sur les travaux réalisés dans le cadre du projet PAMS. Nous développerons un nouveau driver spécifique à Sworm qui assurera l'interconnexion entre l'environnement PAMS et le noyau de simulation Sworm. Ainsi nous profiterons de l'interface web collaborative de PAMS pour contrôler à distance, de manière collaborative, le simulateur Sworm. Grâce aux ressources offertes par le cluster, il sera possible d'exécuter, en parallèle, plusieurs simulations. Nota : ce premier travail d'interconnexion est déjà en cours de réalisation dans le cadre d'un stage de Master 2.

Dans un second temps, nous **modifierons les structures internes du modèle afin de le rendre distribuable**, c'est-à-dire, rendre possible l'exécution d'une même simulation sur plusieurs noeuds de calcul. Le simulateur Sworm présente aujourd'hui des prédispositions à être distribué dans la mesure où il suit les spécifications de l'approche RAFALE-SP, approche dédiée à la création de simulateurs répartis. Mais il n'a jamais été testé dans de telles conditions d'utilisation si bien que nous pouvons imaginer voir apparaître quelques bugs. Nous réaliserons donc une batterie de tests et modifierons le simulateur afin de corriger les bugs apparus.

**WP2: Mise en place d'algorithmes pour la simulation distribuée (Christophe Lang - Nicolas Marilleau) –** ***Besançon/Bondy***

Les stratégies de répartition de charge classiques applicables sur les processus et les objets ont montré leurs limites dans les systèmes multi-agents étant donné la différence dans les caractéristiques des composantes à répartir. En effet, si les processus ont des comportements prévisibles ils ne sont pas forcément faciles à migrer si on ne connaît pas leur état. A l'opposé, à tout instant t, il est possible de connaître l'état d'un agent, on peut les migrer plus facilement. Néanmoins, leur comportement imprévisible pourrait complexifier la répartition de charge : montées en charge imprévisibles, mouvements des agents mobiles sur les machines selon leurs besoins de traitement, etc... Notons également que les relations entre les agents peuvent aussi poser des problèmes dans la mesure où elles risquent d'engendrer une charge de communication supplémentaire.

Dans la littérature, les travaux qui traitent la problématique de la répartition de charge pour les processus classiques sont très nombreux mais ils le sont beaucoup moins pour les plates-formes multi-agents, même si l'on peut trouver quelques références (Auzolat, 2006) (Chow, 2002). De plus, l'objet de notre étude se place dans un sous-domaine encore plus spécifique qui est le domaine des simulations basées sur les systèmes multi- agents. Les spécificités de ces plateformes vont considérablement changer la stratégie de répartition de charge à adopter et les paramètres à prendre en considération lors du processus de migration.

Nous souhaitons dans le cadre de ce projet **mettre en place des routines optimisées de synchronisation (de l'espace et du temps) et de répartition de charge** directement au sein de la bibliothèque RAFALE-SP. Pour cela, nous souhaitons fournir une API permettant à la fois à l'utilisateur d'appeler des fonctions permettant de faciliter la mise en place de telles politiques mais également des politiques de répartition "clés en main". Ainsi, SWORM se basant sur RAFALE-SP nous pourrons appliquer ces algorithmes de manière assez simple.

Les tests se feront sur le méso-centre de calcul de l'Université de Franche-Comté qui est une plate-forme idéale pour une répartition de charge à grande échelle.

Nos travaux se baseront sur deux mémoires de master recherche soutenus en 2009 (Hassoumi, 2009 ; Russo, 2009). Ils donnent une bonne base de travail que cela soit au niveau bibliographie ou au niveau des algorithmes de répartition qu'ils proposent.

**WP3: Optimisation de l'environnement de simulation Madkit pour un déploiement sur cluster (Fabien Michel, Jacques Ferber) –** ***Montpellier***

Les outils de simulation aujourd'hui disponibles dans la plate-forme MadKit (Ferber et al., J., 2004) définissent un ensemble de composants logiciels génériques, basés sur des concepts organisationnels, qui permettent de réaliser des simulateurs multi-agents adaptés à différents domaines d'application (Michel ,2004). Par ailleurs, la plate-forme MadKit possède déjà des fonctionnalités logicielles qui permettent de relier plusieurs instances de la plate-forme en réseau. Ainsi, bien que leur nombre reste restreint, des applications de simulation distribuée ont déjà été développées à l'aide de ces outils. Par exemple, dans un cadre de prototypages académiques (Michel et al. 2002) ou dans le contexte d'un jeu vidéo multi-joueurs en ligne (http://www.jeuxvideopc.com/developpeurs/1143-aleph0/, jeu maintenant abandonné).

Cependant, malgré leur généricité, les outils de simulation la plate-forme MadKit ont principalement été conçus pour définir des simulateurs multi-agents fonctionnant en mono-poste. En effet, le développement d'applications de simulation distribuées nécessitent aujourd'hui la réalisation d'une couche logicielle non triviale qui prend en charge la répartition de la simulation sur les différents noeuds du réseau. En effet, MadKit ne gère pour l'instant que la connexion physique des différentes instances de la plate-forme entre elles. C'est pourquoi, l'ensemble du processus de distribution des différentes entités de la simulation (agents et environnement) reste encore aujourd'hui à la charge du développeur. Dans le contexte actuel, où l'utilisation des ressources mises à disposition sur le réseau est censé devenir de plus en plus naturel, ceci constitue une faiblesse importante de MadKit et, d'une manière générale, un verrou technologique pour lequel il n'existe à l'heure actuelle aucune solution générique.

Ainsi, dans le cadre de ce projet, l'objectif de ce travail repose donc sur une évolution fortement orientée réseau des outils de simulation proposés dans la plate-forme MadKit. En particulier, il s'agira de proposer des solutions logicielles qui permettront de simplifier et d'automatiser la distribution d'une simulation basée sur la plate-forme MadKit. En particulier, la répartition de la charge sur un cluster devra entièrement pouvoir être gérée grâce à des fonctions nativement présentes dans la plate-forme. Cela afin de faciliter les travaux qui seront développés en WP2.

Dans cette perspective, la majeur partie du travail qui sera réalisé se concentrera sur **la définition d'un ensemble de nouvelles fonctionnalités qui seront ajoutées aux outils de simulation existants de MadKit**. Cela de manière à intégrer naturellement la **gestion des aspects liés au réseau et à la distribution sur un cluster**. En particulier, nous nous focaliserons sur des problématiques liées au déploiement à distance d'une simulation distribuée sur un cluster, à son contrôle (ex: récupération des résultats de simulation), et au monitoring (ex. obtention de données système). La première année, ce travail débouchera donc sur une nouvelle bibliothèque d'outils entièrement remaniée qui servira de base aux futures simulations distribuées développées à l'aide de la plate-forme MadKit. La deuxième année sera consacrée aux corrections et modifications qui seront nécessaires en fonction de l'évolution des besoins liés à l'avancement global du projet.

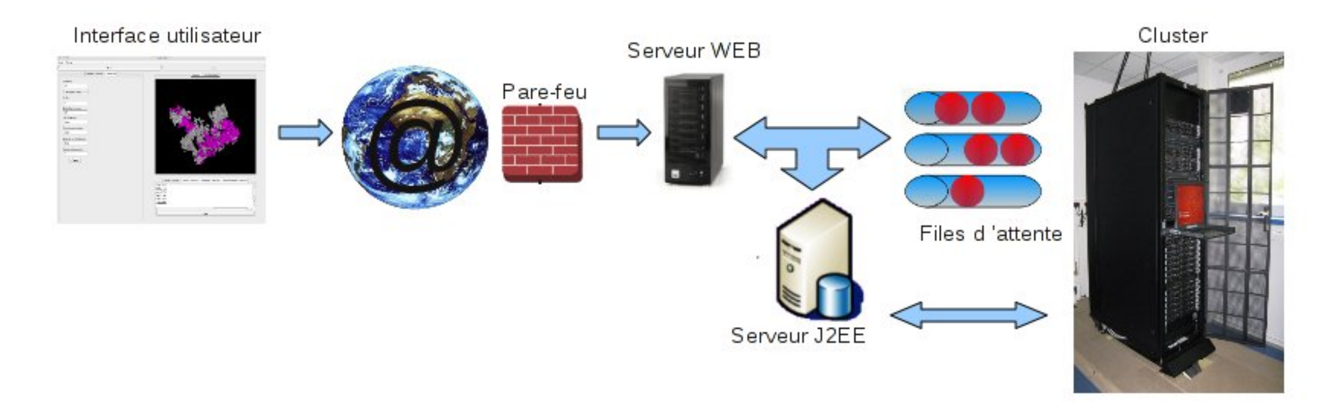
**WP4: Mise en place de routines pour le déploiement et l'exécution de simulateurs sur un cluster de calcul (Laurent Philippe, Clive Canape) –** ***Besançon/Montpellier***

L'exploitation des ressources du cluster suppose une parallélisation du code du simulateur mais aussi une adaptation au mode opératoire en vigueur sur ce type de machine. En effet, les clusters mis à la disposition par les centres de calcul sont des machines partagées par de nombreux utilisateurs. Pour permettre le partage de ces ressources entre les utilisateurs, le système de gestion du clustrer, le batch scheduler, met en place des files d'attente. Un job soumis dans une file d'attente du batch scheduler se voit attribuer une priorité et ce n'est que lorsqu'il est le processus le plus ancien et ayant la priorité adaptée qu'il peut s'exécuter. Or le simulateur est, à l'heure actuelle, prévu pour fonctionner de manière "interactive". Il est donc nécessaire de l'adapter pour permettre l'utilisation des ressources.

Un certains nombre de défis se posent donc quant à l'utilisation d'un cluster pour l'exécution du simulateur. Le premier est la mise en place d'un environnement d'exécution adapté à la version parallèle. En effet, la soumission sous forme de job d'une instance de simulation suppose que le simulateur, au moment où il est exécuté à distance par

le bacth scheduler, retrouve bien tout l'environnement nécessaire à son exécution : classes, fichiers, accès aux bases, etc. Le second aspect qui nous semble mériter une attention particulière est la connexion avec la base de données. Il sera nécessaire de gérer les synchronisations d'accès pour garantir sa cohérence, de valider que son implémentation centralisée lui permet quand même de supporter le passage à l'échelle et de vérifier que l'interaction avec le mode batch ne pose pas de problème. Enfin le mode opératoire avec soumission non interactive n'est pas une utilisation classique d'un simulateur. **Il conviendra donc de mettre en place et automatiser les procédures de distribution et de lancement des instances**. Une attention tout particulière doit être portée à la portabillité de ces procédures. En effet elles n'auraient que peu d'intérêt si elles ne permettaient que l'exécution sur le cluster de test.

Pour faciliter l'accès des chercheurs non-initiées, auquel il est difficile de demander cette soumission à travers un batch scheduler, il est également nécéssaire de fournir une interface facilitant le lancement d'une instance de simulation. Il s'agira d'une interface web permettant de saisir les paramètres et de préciser l'environnement d'exécution. Ce type daccès permet également de cacher les complexités d'accès au cluster, protégé derrière un pare-feu.




Architecture d'accès au simulateur Sworm

Comme cela est montré sur la figure, l'architecture envisagée reposera sur un serveur web, accessible depuis l'extérieur, qui interagira avec un serveur d'application et le cluster. Le rôle du serveur d'application est de gèrer l'accès à une base de donnée, de manière à stocker les environnements et les résultats de simulation, et d'exécuter les simulations.

**WP5 Conception d'interfaces web didactique pour la programmation de plans d'expériences, le contrôle et l'analyse de résultat (Tuong Vinh Ho, Benoit Gaudou, Eric Blanchart, Jean Claude Kamgang)**

Le rôle principal de notre interface est de permettre à l'utilisateur, de manière simple et intuitive, de paramétrer et de définir les plans d'expériences, de lancer les simulations et de permettre une analyse poussée des résultats. Nous souhaitons ainsi cacher toute la complexité de l'exécution d'un simulateur sur un cluster ou une grille en particulier pour des thématiciens qui sont les demandeurs et les futurs utilisateurs de notre application. Compte-tenu du public visé, une interface web nous apparaît la plus adaptée.

Les avantages d'une interface web sont multiples : (i) elle ne nécessite aucune installation de la part de l'utilisateur car elle n'a besoin que d'un navigateur web, application aujourd'hui disponible sur l'immense majorité des ordinateurs ; (ii) l'utilisateur est en mesure de conserver son environnement de travail même s'il change de station de travail, de système d'exploitation ; (iii) elle ne nécessite aucune maintenance, ni configuration de la part de l'utilisateur. Sa principale contrainte est liée à la nécessité d'une connexion réseau ou internet avec une bande passante suffisante.

L'enjeu de ce WP est donc de **construire une interface intuitive**, adaptée (aux coutumes et usages des thématiciens), et **optimisée** afin qu'elle puissent être utilisée dans des conditions «extrêmes» où le réseau serait dégradé comme cela peut être constaté dans les pays du Sud. Pour atteindre ses buts, notre interface sera construite en collaboration avec des thématiciens (biologistes et mathématiciens), futurs utilisateurs, et devra être pourvue de fonctionnalités permettant l'initialisation des simulation, le contrôle des simulations et l'analyse des données de résultat.

**L'interface d'initialisation des simulations** vise à aider le paramétrage des expérimentations avant leurs réalisations effectives sur le cluster du mésocentre. Autrement dit, il s'agira d'un outil d'assistance à la définition de plan d'expériences. Cette interface permettra d'établir automatiquement un ensemble de simulations à réaliser en fonction d'un espace de paramètres établi par l'utilisateur, où certains paramètres sont fixes (ex. la granulométrie du sol, et d'autres sont définis par des bornes inférieures et supérieures (ex. le nombre de vers de terre simulés). Par ailleurs, l'interface proposera également un historique des plans d'expériences proches, des résultats de simulation qu'ils ont permis de produire, et des commentaires éventuels déposés. Ainsi, le thématicien sera en mesure de capitaliser des expérimentations déjà réalisées pour s'en inspirer et proposer de nouvelles expérimentations spécifiques à son cas d'étude. Des fonctionnalités avancées de paramétrage du cluster pourront aussi être proposées par l'interface, par exemple pour imposer un ordonnancement et une parallélisation fins des simulations. Leur utilisation ne sera pas obligatoire car l'environnement que nous proposons sera pourvu de fonctions automatiques de parallélisation mais pourra dans certains cas s'avérer nécessaire.

**L'interface de contrôle** sera un moyen pour l'utilisateur de suivre en temps réel le déroulement des simulations. Il s'agira d'une interface de synthèse et en aucun cas d'une interface d'analyse. Autrement dit, elle renseignera l'utilisateur sur l'état des expérimentations (ex. temps restant avant la fin des simulations et quelques resultats intermédiaires) afin de valider ou d'invalider le choix des paramètres. Grâce à ces informations, le chercheur sera en mesure de prendre connaissance d'éventuels problèmes et donc d'interrompre les simulations, de modifier le paramétrage et de relancer les simulations (soit depuis le début, soit à partir de points d'arrêt). L'affichage de cette interface sur des clients légers tels que des téléphones portables est envisageable.

**L'interface d'analyse** vise à assurer un traitement avancé des résultats. Deux possibilités seront proposées à l'utilisateur : soit le chercheur exporte les résultats bruts de simulation pour les analyser via son outil statistique préféré ; soit le chercheur utilise les outils d'analyse et de visualisation proposés par l'outil. Cette dernière peut s'avérer très utile pour réaliser un contrôle rapide, toutefois avancé, des simulations avant un export des données. Par ailleurs, nous souhaitons coupler cette interface avec un outil collaboratif comme PAMS. Ainsi, elle aura l'originalité et l'intérêt de :

- **favoriser la coopération** entre plusieurs chercheurs dispersés de part le monde, au Nord et au Sud, de l'analyse des résultats ;
- **permettre l'annotation et la discussion des résultats** obtenus par l'ensemble des chercheurs de la communauté.

Bien que dans un premier temps, ce projet et les outils développés s'organisent autour du modèle Sworm, nous nous efforcerons à ce que ces interfaces soient le plus génériques et modulaires possible afin de tendre vers en environnement web de paramétrage, de contrôle et de traitement de lots de simulations. Leur construction nécessitera une étroite collaboration entre informaticiens et thématiciens (dans un premier temps biologiste, ensuite épidémiologiste, géographe…). Ces derniers, futurs utilisateurs de l'outil, sont parties prenante dans le projet.

**WP6: Formation (Nicolas Marilleau, Eric Blanchart) *Bondy***

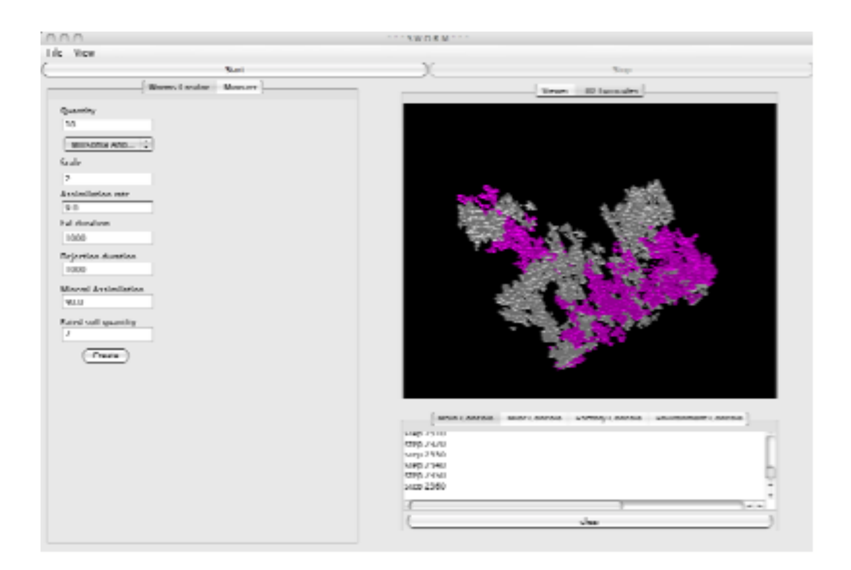
A l'issue de ce projet, **une formation utilisateurs sera organisée dans un premier temps à Bondy**. Le public visé sera constitué de thématiciens et de modélisateurs en science du sol. Lors de cette formation, un tutoriel suivit de **démonstrations** seront réalisés. Dans un second volet, cette formation iniciera à l'intégration de nouveaux modèles dans Epis

## 5.3 Description de l'existant (moyens – outils – compétences)

*OU renvoyer à un document joint*

### 5.3.a Simulateur Sworm

**Le simulateur Sworm (*Marilleau et al. 2008 ; Blanchart et al. 2009*)** fait suite au projet ANR Microbes (Chotte, 2006-2009) en sciences du sol et est le résultat d'une collaboration entre deux unités de l'IRD : Seq-Bio (UMR Eco&Sol) et Geodes (UMI UMMISCO). Ce simulateur vise à reproduire la dynamique de la macrofaune des sols tropicaux. Ce modèle permet entre autres de suivre l'organisation spatiale des déjections de vers de terre dans le sol ainsi que la dynamique de la matière organique. Les résultats obtenus par ce nouveau modèle sont novateurs et présente un intérêt important pour la communauté de sciences du sol. L'article (Blanchart et al., 2009) était en 2009 l'un des 10 articles les plus lus de la revue *European Journal in Soil Sciences,* revue qui fait référence en sciences du sol.



*Capture d'écran du simulateur Sworm*

Le simulateur Sworm est aujourd'hui standalone. Il a été développé en Java en étendant la bibliothèque RAFALE-SP qui, elle-même, repose sur l'environnement de simulation Madkit. Aujourd'hui, Sworm n'est pas un simulateur réparti mais sa distribution se fera à moindre coût. Il suit les spécifications de RAFALE-SP ce qui lui confère des prédispositions à la distribution.

### 5.3.b PAMS (http://www.pamsproject.org)

La plate-forme PAMS (Nguyen *et al.*, 2009) est une première initiative menée dans le domaine de la modélisation-simulation et financée par le programme Spirales de l'IRD (2007-2008). Ce collecticiel permet, à distance, via une interface web collaborative (voir ci-dessous) : la manipulation et l'exécution coopératives de simulateurs, le partage des expérimentations et la capitalisation des résultats.

PAMS combine des outils collaboratifs généralistes (vidéo-conférence, messagerie instantanée, etc) avec des outils collaboratifs spécifiques au domaine des systèmes complexes (ex. panneaux collaboratifs de contrôle des simulateurs, visualiseur de résultats, etc). Ces derniers ont spécifiquement été développés dans le cadre du projet Spirales PAMS.




*Capture d'écran du portail Pams*

L'environnement PAMS est un système distribué pouvant par définition être déployé sur un cluster d'ordinateurs. Cette plate-forme se compose :

- **d'un serveur WEB** (Tomcat) pour la gestion de l'interface utilisateur ;
- **d'un serveur d'applications J2EE** (Jonas) pour la gestion de l'application (exécution des simulateurs, traitement des données, …) ;
- **d'une base de données** (MySQL) pour le stockage des modèles, des expérimentations et des résultats.

Ce collecticiel repose donc sur une architecture orientée « composant » lui conférant la généricité, la modularité et l'extensibilité nécessaires pour répondre aux besoins actuels et de demain, des chercheurs. Cette généricité permet d'intégrer au sein de PAMS des plates-formes de simulation comme NetLogo et Repast. Il est aussi possible d'installer des simulateurs développés de manière ad-hoc.

La version actuelle de PAMS supporte nativement les simulateurs développés à partir des plates-formes NetLogo, GAMA et Repast. Nous l'avons testée en déployant des simulateurs « jouets » dont le célèbre « jeu de la vie ». Par ailleurs, deux simulateurs faisant l'objet d'une recherche active menée par quelques thématiciens de l'IRD, ont été installés : l'un, Gamavi, s'inscrit dans le domaine de l'épidémiologie, l'autre, Simyfor, entre dans le cadre de recherche en sciences du sol.

Pour davantage d'informations sur l'environnement PAMS, le lecteur peut se référer aux délivrables du projet, au manuel d'utilisation et au support de formation.

### 5.3.c RAFALE-SP

Au terme du projet Miro, nous avons réussi à proposer une extension de Madkit, appelée RAFALE-SP (Marilleau, 2006), qui a été éprouvée avec succès dans des thématiques de recherches très différentes dont la géographie et les sciences du sol (Marilleau *et al.*, 2008).

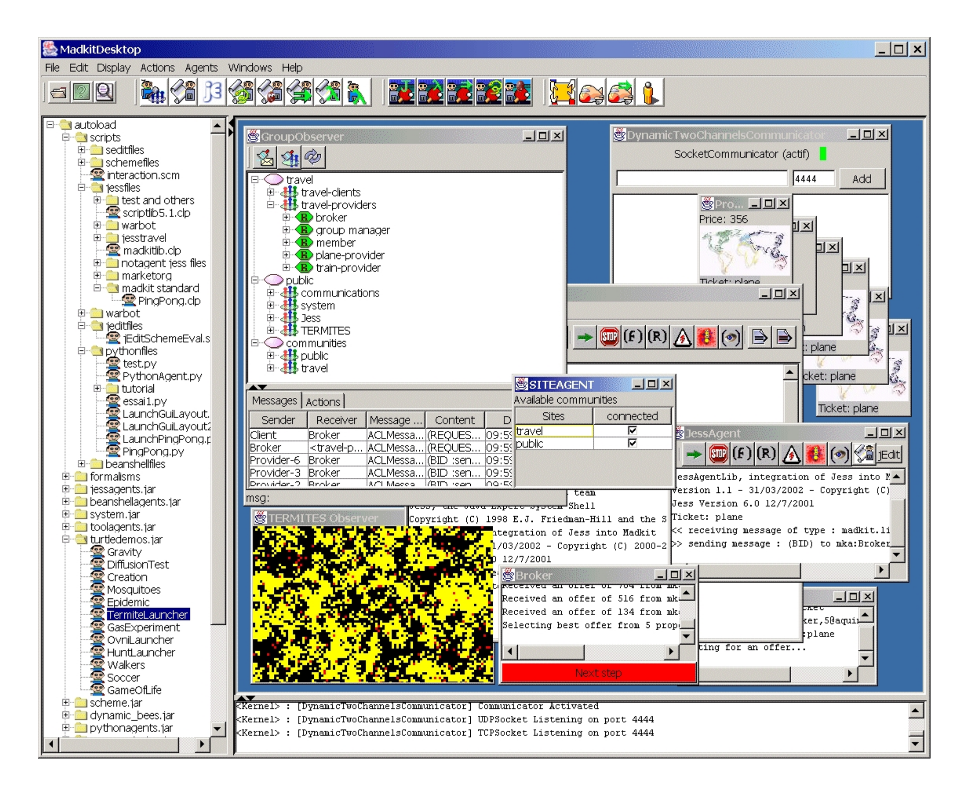
Ce nouvel environnement présente l'avantage d'être générique si bien que le modélisateur est libre de choisir et d'implémenter la structure la plus adaptée pour modéliser un espace et des individus. RAFALE-SP reste néanmoins une plate-forme de simulation des systèmes complexes et dispose donc de tous les éléments conceptuels et logiciels requis aidant le modélisateur dans le développement de ses simulateurs. Elle intègre par exemple tout un ensemble

de routines qui permettent de distribuer un simulateur sur une grille et ainsi d''opérer des simulations à large échelle. Dans le cadre du précédent projet MIRO, nous avons eu l'occasion d'évaluer les performances de la plate-forme, avec différentes configurations, au travers d'expérimentations qui consistaient à faire évoluer dans un environnement virtuel représentatif de la ville de Dijon, des agents browniens qui choisissent aléatoirement leur trajet tout en respectant les règles de circulation.

Toutefois, RAFALE-SP est davantage un toolkit qu'une réelle plate-forme de simulation. La création d'un simulateur avec cet outil nécessite une phase de développement (en java) importante. Par conséquent, RAFALE-SP ne peut pas être utilisé par un thématicien ou un modélisateur dénué de compétences fortes en développement informatique. La généricité de RAFALE-SP permet malgré tout la création d'une interface graphique adaptée aux besoins du thématicien et du projet. Par conséquent, même si les thématiciens ne peuvent pas intervenir lors de la phase de développement, ils sont en mesure d'utiliser, par leurs propres moyens, le simulateur résultant, et donc de participer activement à la validation du modèle et des simulateurs implémentés et aux expérimentations. Le développement de l'interface d'un simulateur s'effectue au cas par cas car RAFALE-SP ne dispose que de très peu de primitives prévues à cet effet. Par ailleurs, cet outil reste perfectible autant d'un point de vue conceptuel (certains concepts de modélisation peuvent être améliorés notamment en ce qui concerne la représentation des connaissances et la validation des modèles) que d'un point de vue opérationnel (notamment en ce qui concerne les algorithmes de parallélisation et de répartition de charge).

### 5.3.d MadKit (http://www.madkit.org)

MadKit est une plate-forme de développement de systèmes multi-agents destinée au développement et à l'exécution de systèmes multi-agents et plus particulièrement à des systèmes multi-agents fondés sur des critères organisationnels (groupes et rôles).




*Capture d'écran de la plate-forme MadKit*

Madkit fonctionne de manière distribuée selon un mode peer to peer. Il introduit aussi la notion de « communauté », c'est à dire d'application distribuée à laquelle un ensemble d'agents peuvent se réunir et partager des applications et des documents.

MadKit n'impose aucune architecture particulière aux agents. Il est ainsi possible de développer aussi bien des applications avec des agents réactifs (comme le fait TurtleKit) que des agents cognitifs et communicationnels, et même de faire interagir aisément tous ces types d'agents. Cela permet ainsi aux développeurs d'implémenter l'architecture de leur choix.

MadKit est écrit en Java et fonctionne en mode distribué de manière transparente à partir d'une architecture "peer to peer" sans nécessiter de serveur dédié. Il est ainsi possible de faire communiquer des agents à distance sans avoir à se préoccuper des problèmes de communication qui sont gérés par la plate-forme.

Le mécanisme de distribution est indépendant de MadKit et il est possible d'en créer de nouveaux si les besoins s'en font sentir ou même d'utiliser des plate-formes distribuées existantes, telles que JXTA. MadKit est construit autour du concept de "micro-noyau" et "d'agentification de services". Le micro-noyau de MadKit est petit (environ 100Ko de code), car il ne gère que les organisations (groupes et rôles) et les communications à l'intérieur de la plate-forme. Le mécanisme de distribution, les outils de développement et de surveillance des agents, les outils de visualisation et de présentation sont des outils supplémentaires qui viennent sous la forme d'agents que l'on peut ajouter au noyau de base.

De ce fait, MadKit peut être utilisé aussi bien comme outil de développement d'applications que comme un noyau d'exécution de système multi-agent pouvant être embarqué dans des applications quelconques.

MadKit est un logiciel libre de type "Open Source" avec une licence mixte GPL/LGPL. LGPL pour le micro-noyau et les outils de communication, et GPL pour les outils de développement.

On reconnait généralement à MadKit les qualités suivantes :

- **Simplicité** de mise en oeuvre et de prise en main,
- **Intégration** très facile de la distribution au sein d'un réseau.
- **L'aspect pratique et efficace des concepts organisationnels** pour créer différents types d'applications
- **Hétérogénéité** des applications et des types d'agents utilisables: on peut faire tourner sur MadKit aussi bien des applications utilisant des agents réactifs simples de type fourmis (plusieurs dizaine de milliers d'agents ayant un comportement simple ont tourné en MadKit), que des applications disposant d'agents cognitifs sophistiqués.

### 5.3.e Le méso-centre de calcul de l'Université de Franche-Comté

Depuis le mois de novembre 2009, l'Université de Franche-Comté dispose d'un méso-centre de calcul équipé d'un cluster, ensemble d'ordinateurs interconnectés par un réseau rapide, d'une puissance crète de 5 Tflpos et de logiciels qui peuvent être utilisés à la demande. Le but de ce mésocentre est de proposer au niveau de l'université et de la région des ressources de calcul haute-performance et de développer l'expertise locale dans le domaine de l'utilisation de ce type de ressources. Le personnel du mésocentre est constitué d'un ingénieur de recherche, d'un ingénieur d'étude et d'un assistant ingénieur. Ils administrent les ressources et fournissent une aide aux utilisateurs non expérimentés pour leur faciliter l'accès au calcul haute performance. D'un point de vue technique, la configuration acquise par le mésocenttre est un cluster Bullx constitué de 68 noeuds de calcul, chacun équipé de deux processeurs (les nouveaux processeurs Nehalem d'Intel) quad-coeurs ce qui fait un total de 544 coeurs. Chaque noeud dipose de 12 à 24 Go de mémoire centrale, un noeud dispose de 96 Go. Les noeuds sont interconnectés par un réseau InfiniBand d'un débit de 20 Go/s, destiné à assurer des communications rapides entre les programmes, et d'un réseau gigaethernet pour les échanges courants. La capacité de stockage associée au cluster est de 20 To gérés en Raid 6 depuis un noeud spécifique et partagés par NFS sur les noeuds de calcul, plus un espace de sauvegarde de 10 To. L'ensemble est accédé à travers trois noeuds de login, permettant la compilation, la mise au point et la soumission des programmes. L'attribution des noeuds de calcul est assurée par le gestionnaire de file d'attente Sun Grid Engine (SGE). L'administration est réalisée depuis un noeud spécifique. L'ensemble est intégré à deux armoires réfrigérées qui limitent l'impact environnemental en optimisant la consommation énergétique due à la climatisation.



*Cluster mis en place à l'Université de Franche-Comté*

Pour exploiter ces ressources le mésocentre de calcul met à disposition de ses utilisateurs un grand nombre de logiciels permettant aussi bien la réalisation de programme parallèles ou l'utilisation de bibliothèques spécialement optimisées que l'exécution de code de calcul déjà existants. Parmi les logiciels les plus couramment utilisés nous pouvons citer l'Intel Cluster Toolkit, Matlab, VASP, Gaussian, NAMD, etc. Les logiciels payants mis à disposition le sont pour le moment avec des licences académiques et sont donc réservés aux chercheurs ou enseignants mais le mésocentre dispose d'un grand nombre de logiciels issus du monde de l'open source qui peuvent parfois remplacer l'utilisation d'un logiciel propriétaire.

### *5.4 Difficultés rencontrées jusqu'à présent :*

De nombreuses difficultés sont constatées par les thématiciens quant à l'utilisation de simulateurs dont :

- La complexité d'utilisation et d'installation des simulateurs
- L'impossibilité de realiser des simulations nécessitant une forte puissance de calcul
- La difficulté à communiquer autour du modèle Sworm

## 5.5 Références


(Auzolat, 2006) : Arnaud Auzolat, 2006: Modèle organisationnel de répartition de charges pour des agents dans des contextes contraints. Application à la résolution de problèmes et à la simulation multi- agent distribuée sur le GRID, Université Montpellier II.

(Blanchart *et al.* 2009) E. Blanchart, N. Marilleau, J-L. Chotte, A. Drogoul, E. Perrier and Ch. Cambier. SWORM: an agent-based model to simulate the effect of earthworms on soil structure. *European Journal of Soil Science*. 2009;60(1):13-21. Available at: http://blackwell-synergy.com/doi/abs/10.1111/j.1365-2389.2008.01091.x.

(Chow *et al.*, 2002) Ka-Po Chow and Yu-Kwong Kwok, Member, IEEE : On Load Balancing for Distributed Multiagent Computing

(Ferber et al., J., 2004) J. Ferber, O. Gutknecht and F. Michel, "From Agents to Organizations: an Organizational View of Multi-Agent Systems" in Agent-Oriented Software Engineering IV, Paolo Giorgini and Jörg P. Müller and James Odell eds., pp 214—230, LNCS 2935, 2004.

(Hassoumi, 2009) *Mise en place et évaluation d'un algorithme de répartition de charge pour les plates-formes de simulation distribuées basées sur les systèmes multi-agents,* Hassoumi Inès, mémoire de master recherche, Université de Franche-Comté, Aout 2009.

(Marilleau *et al.* 2008) N. Marilleau, Ch. Cambier, A. Drogoul, E. Perrier, J-L. Chotte and E. Blanchart Multiscale MAS modelling to simulate the soil environment: Application to soil ecology. *Simulation Modelling Practice and Theory*. 2008;16(7):736-745.

(Marilleau and al., 2007) N. Marilleau, C. Cambier, A. Drogoul, E. Perrier, JL. Chotte and E. Blanchart. Multiscale MAS modeling to simulate complex systems: A case study in soil science. In MACS2007, Multi-agents for modelling complex systems, Dresden, Allemagne, 2007.

(Marilleau, 2006) N. Marilleau, Méthodologie, formalismes et outils de modélisation-simulation pour l'étude des systèmes complexes : application à la mobilité géographique, Thèse de doctorat, Université de franche-Comté, Besançon, novembre 2006

(Michel ,2004) Michel Fabien, "Formalisme, outils et \'el\'ements m\'ethodologiques pour la mod\'elisation et la simulation multi-agents", thèse de doctorat en informatique, Soutenue le 21 décembre 2004 à l'université de Montpellier II.

(Michel et al. 2002) F. Michel, Pierre Bommel et J. Ferber, Simulation interactive distribuée sous MadKit, JFIADSMA'02, 10èmes journées Francophones pour l'Intelligence Artificielle Distribuée et les Systèmes Multi-Agents (Lille 2002) Actes des JFIADSMA'02, Hermès, p. 175-178

(Nguyen *et al.*, 2009) T-K. Nguyen, B. Gaudou, T-V. Ho and N. Marilleau *Application of PAMS Collaboration Platform to Simulation-Based Researches in Soil Science: The Case of the MIcro-ORganism Project.* Dans : *IEEE International Conference on Research, Innovation and Vision for the Futur (RIVF 2009)*, *Da Nang City, Viet Nam*, 2009.

(Russo, 2009) *Contribution à l'élaboration d'un algorithme de répartition de charge par migration de l'environnement pour les simulations multi-agents,* Russo David, mémoire de master recherche, Université de Franche-Comté, Aout 2009.

(Sebastien 2009) N. Sebastien, Distribution et Parallélisation de Simulations Orientées Agent, thèse de doctorat, Université Saint Denis de la Réunion, soutenue le 10 novembre 2009.

## 6. Description du projet – SEULEMENT SI « développement d'application IS » (méthodes, solutions, et moyens)

La demande peut-être être **accompagnée de tous documents utiles** :

*présentation du projet global ou descriptif du projet, rapport de phases*

*préliminaires, étude de faisabilité (définition des besoins), dossier d'expression des besoins ou **cahier des charges**, devis détaillé…*

### ***6.1. Nom de votre outil***

EPIS : Easy, Parallel and Interactive Sworm

### ***6.2. Si votre outil existe déjà, quel est l'URL du site internet ou des documents qui le décrivent? Ou, si l'outil a été décrit dans un article, fournir les références***

---

## ***1. Innovation :***

### ***6.3. Ecrire 3 scénarios qui illustrent comment votre outil sera ou a été utilisé dans votre communauté scientifique ou domaine d'activités***

**Scénario 1 : *« Expérimentation collaborative à distance »***
X et Y, chercheurs en sciences du sol, respectivement localisés à Bondy et à Hanoi, souhaitent réaliser quelques expérimentations Sworm portant sur le territoire de Lamto (Côte d'Ivoire). Ces expérimentations visent à illustrer l'article qu'ils écrivent en commun. Les deux chercheurs se connectent sur le site web du projet et lancent ensemble une discussion privée et l'interface de configuration du simulateur Sworm. M. X commence à saisir des paramètres de simulation, qui ne conviennent pas à M . Y, lequel démarre une session de visioconférence pour faire part à M. X de sa surprise. Il pose également sur l'interface du simulateur une note indiquant la valeur des paramètres qui lui semble correcte. Après discussion, et affichage, par M. Y, de la partie d'article qu'ils souhaitent illustrer, les deux chercheurs décident de réaliser les deux expérimentations et de les comparer ensuite. A l'aide de l'interface, X et Y construisent ensemble un plan d'expérience incorporant ces deux expérimentations et le valide. Un message d'information apparaît alors indiquant : « compte tenu de la charge actuelle du système, il nous est impossible de traiter votre demande immédiatement. Vos résultats seront disponibles à partir de 18 h 30. Vous serez néanmoins averti par courriel lorsque votre demande sera traitée ». X et Y ayant pris connaissance du message décident d'interrompre leur réunion jusqu'au lendemain.

**Scénario 2 : *« Etude paramétrique »***

M. X, chercheur en sciences du sol, souhaite réaliser une étude paramétrique du modèle Sworm appliqué au sol de Lamto (Côte d'ivoire). M. X se connecte alors sur le portail du projet et ensuite à l'interface de simulation du simulateur Sworm. A partir de cette interface, M. X saisit : (i) les caractéristiques du sol de Lamto dont sa granulométrie ; (ii) et l'espace des paramètres du simulateur. Après cette première étape, M. X clique sur le bouton "Start". Le simulateur est alors chargé et lancé sur plusieurs dizaines de noeuds du clusters. Des informations sur la disponibilité des résultats sont alors affichées. M. X se déconnecte. Quelques heures plus tard et après 5 heures de calcul, M. X reçoit un mail lui notifiant l'achèvement de son expérimentation. M. X se connecte donc de nouveau au portail (via un lien présent dans son mail), affiche les résultats et réalise une analyse avancée en utilisant les outils statistiques proposés par le portail. A partir de cette analyse, M. X génère des graphiques et des données qu'il exporte sous la forme de fichiers images et de fichiers CSV.

**Scénario 3 : *« Contrôle des simulation en cours »***

M. X, chercheurs en science du sol, a programmé, il y a quelques heures, un plan d'expériences qui nécessite la réalisation de 48 simulations et environ 72 heures de calcul. Afin de s'assurer du bon déroulement de son expérimentation, M. X souhaite réaliser des contrôles intermédiaires. Pour cela, il se connecte via son ordinateur (ou, à défaut d'infrastructure, via son "smartphone") au portail et, ensuite, à la simulation. Lorsque M. X est

connecté, une première information lui indique l'état d'avancement (% des simulations réalisées) du plan d'expériences, le temps écoulé, et le temps restant estimé. M. X souhaite s'assurer de la pertinence des paramètres choisis dans le plan d'expériences. Il choisit au hasard, dans l'interface web une simulation réalisée et visualise brièvement les résultats. Ces derniers ne répondant pas aux attentes du chercheur, M. X décide donc d'arrêter le calcul et de modifier le plan d'expérience en ajoutant de nouvelles simulations à réaliser. Après modification du plan d'expériences, M. X relance le calcul et se déconnecte du portail. Les simulations reprennent alors là où elles se sont arrêtées. Le chercheur se déconnecte du portail.

### 6.4. Décrire, en un paragraphe, les innovations de votre projet pour votre communauté scientifique

Les innovations de ce projet sont multiples dont :

- la mise en place outils didactiques pour l'exécution de simulateurs sur un cluster
- la création d'algorithmes permettant la réalisation de simulations réparties (1 expérimentation = n noeud) et/ou parallèle (1 expérimentation = 1 noeud)
- la conception d'outils avancés orientés web pour la mise en place de plan d'expériences
- la conception d'outils avancés orientés web pour l'analyse de résultat de simulation.

### 6.5. Existent-ils d'autres outils similaires au vôtre ? Si c'est le cas, lister ces outils et décrire les avantages de votre outil par rapport aux autres

La problématique abordée dans ce projet est un domaine de recherche à part entière qui est jeune et d'actualité. De nombreux travaux existent : certains s'intéressent l'exécution de simulateurs sur des serveurs dédiés à partir d'une interface web, d'autres tentent de réaliser des simulations dans un environnement réparti. Mais, à notre connaissance, aucun de ces travaux est transversal au domaine de la collaboration, du parallélisme et de la simulation.

### 6.6. Si vous proposez des améliorations à un outil existant, combien d'utilisateurs ont déjà téléchargés ou obtenus une copie de la version actuelle ?

Le simulateur Sworm ne peut pas être téléchargé dans la mesure où il nécessite le concours de ses concepteurs pour sont installation et son utilisation. Par contre, nous comptons aujourd'hui plus d'une trentaine d'utilisateurs de PAMS. Ce groupe est constitué d'informaticiens, de mathématiciens, de thématiciens. Ces chercheurs seront, de manière transparente, utilisateurs de l'outil visé par ce projet .

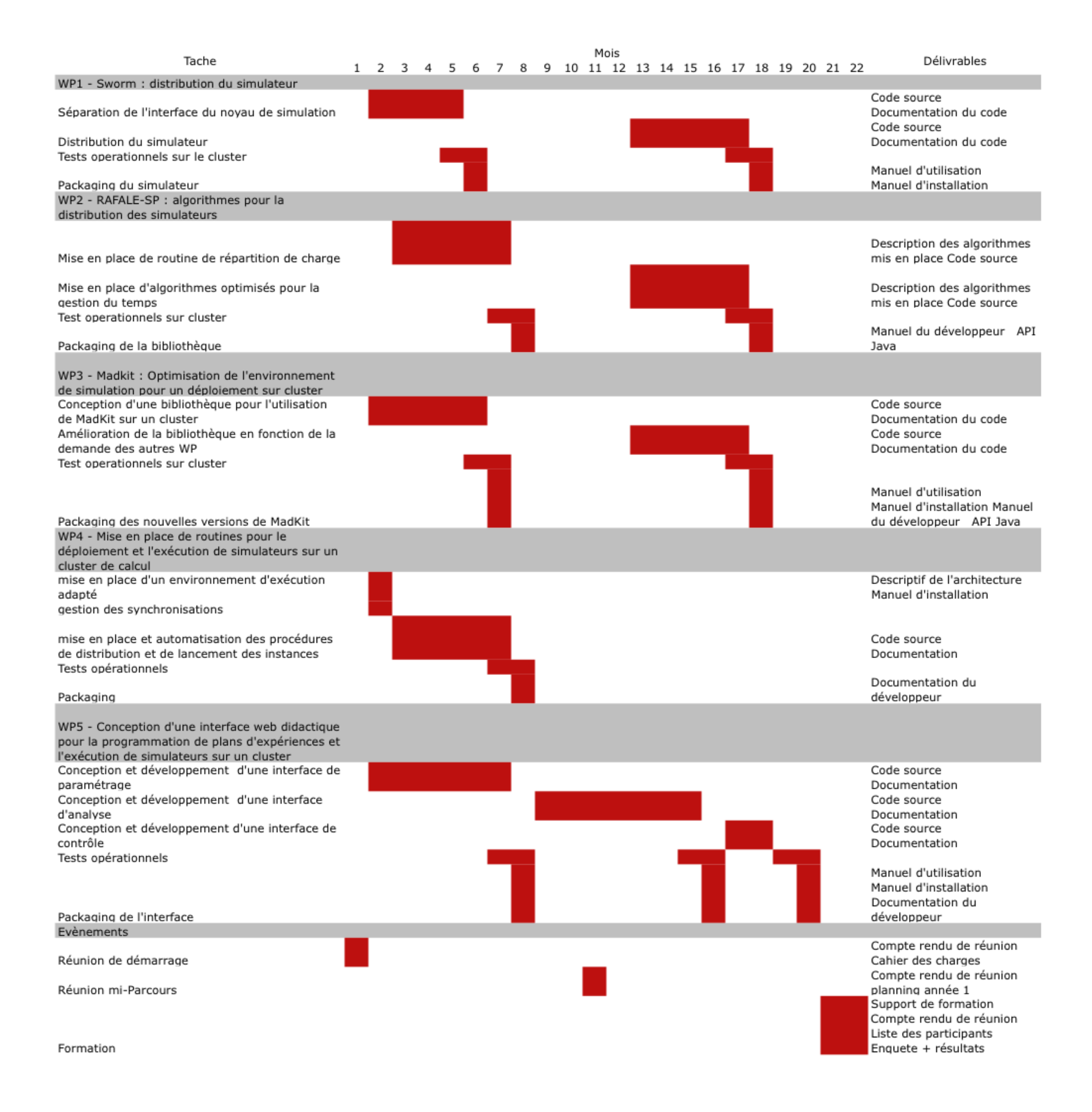
### 6.7. Le projet proposé est-il basé sur de nouvelles conclusions scientifiques ou méthodes innovantes ? Si c'est le cas, décrire les fondements et lister les références les plus pertinentes.

Ce projet est basé sur plusieurs conclusions scientifiques innovantes :

- Le modèle Sworm, est en soi une innovation dans la mesure où il s'agit d'un des premiers modèles multi-agents de ce type mise en place pour répondre à des problèmes de sciences du sol. Les résultats obtenus sous tend des perspectives intéressantes.
- Les travaux menés autour de la bibliothèque RAFALE-SP (présentés dans Marilleau 2006) ont montré qu'une distribution adaptée des simulateurs pouvait entraîner une réduction par n du temps de calcul (n étant le nombre de noeuds). Par exemple, si une simulation est répartie sur 2 noeuds, nous pouvons espérer diviser par 2 le temps de simulation par rapport à une même simulation non répartie.

## 2. Calendrier, budget et risques

### 6.8. Calendrier du projet montrant les tâches clés et les dates d'échéances



Tache
Mois
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Délivrables
WP1 - Sworm : distribution du simulateur
Séparation de l'interface du noyau de simulation
Code source
Documentation du code
Distribution du simulateur
Code source
Documentation du code
Tests operationnels sur le cluster
Packaging du simulateur
Manuel d'utilisation
Manuel d'installation
WP2 - RAFALE-SP : algorithmes pour la distribution des simulateurs
Mise en place de routine de répartition de charge
Description des algorithmes mis en place Code source
Mise en place d'algorithmes optimisés pour la gestion du temps
Description des algorithmes mis en place Code source
Test operationnels sur cluster
Packaging de la bibliothèque
Manuel du développeur API Java
WP3 - Madkit : Optimisation de l'environnement de simulation pour un déploiement sur cluster
Conception d'une bibliothèque pour l'utilisation de MadKit sur un cluster
Code source
Documentation du code
Amélioration de la bibliothèque en fonction de la demande des autres WP
Code source
Documentation du code
Test operationnels sur cluster
Packaging des nouvelles versions de MadKit
Manuel d'utilisation
Manuel d'installation Manuel du développeur API Java
WP4 - Mise en place de routines pour le déploiement et l'exécution de simulateurs sur un cluster de calcul
mise en place d'un environnement d'exécution adapté
Descriptif de l'architecture
Manuel d'installation
gestion des synchronisations
mise en place et automatisation des procédures de distribution et de lancement des instances
Code source
Documentation
Tests opérationnels
Packaging
Documentation du développeur
WP5 - Conception d'une interface web didactique pour la programmation de plans d'expériences et l'exécution de simulateurs sur un cluster
Conception et développement d'une interface de paramétrage
Code source
Documentation
Conception et développement d'une interface d'analyse
Code source
Documentation
Conception et développement d'une interface de contrôle
Code source
Documentation
Tests opérationnels
Packaging de l'interface
Manuel d'utilisation
Manuel d'installation
Documentation du développeur
Evènements
Réunion de démarrage
Compte rendu de réunion
Cahier des charges
Réunion mi-Parcours
Compte rendu de réunion
planning année 1
Formation
Support de formation
Compte rendu de réunion
Liste des participants
Enquete + résultats

### ***6.9. Eventuellement, budget détaillé montrant les coûts des tâches clés, des différents modules ou phases.***

(*Les informations apportées doivent être cohérentes avec celles précisées à la question 18.*)

| Tache | Objet | année 1 | année 2 |
|---|---|---|---|
| WP1 - Sworm : parallélisation du simulateur | 1 stage / an 6 mois 400€ / mois | 2400 | 2400 |
| WP2 - RAFALE-SP : algorithmes pour la distribution des simulateurs | 1 stage de 6 mois 400€ / mois | 0 | 2400 |
| WP3 - Madkit : Optimisation de l'environnement de simulation pour un déploiement sur cluster | 1 stage / an 6 mois 400€ / mois | 2400 | 2400 |
| WP4: Mise en place de routines pour le déploiement et l'exécution de simulateurs sur un cluster de calcul | 1 serveur (2000 €) 1 stage de 6 mois 400€ / mois | 4400 | 0 |
| WP5 Conception d'une interface web didactique pour la programmation de plans d'expériences et l'exécution de simulateurs sur un cluster | 18 mois 200 €/mois | 1800 | 1800 |
| WP6 Formation | Restauration + Fourniture + Déplacements des formateurs | | 2500 |
| Déplacements | France ←→Vietnam + France←→France | 4000 | 4000 |
| | | | |
| | **Total** | 15000 | 15500 |
| | | | |
| | | | |
| | **Fonctionnement** | 4000 | 6500 |
| | **Prestation de service** | 9000 | 9000 |
| | **Equipement** | 2000 | 0 |

### ***6.10. Si vous demandez des fonds pour des activités autres que du « développement logiciel », pourquoi ces activités sont-elles essentielles à l'accomplissement de votre projet ?***

Le cluster qui est en possession de l'Université de Franche Comté ne dispose pas de connexion vers le réseau internet. Pour le succès de ce projet, il est donc important de mettre en place un serveur de tête qui fera l'interface entre Internet et le réseau privé du méso centre.

Le projet fait appel à des partenaires appartenant à 4 unité différentes (UMMISCO, LIFC, LIRMM et DSI). Pour cette raison nous allouons un budget spécifique pour le déplacement afin de réaliser les réunions de démarrage et intermédiaires.

### ***6.11. Quel sont les risques encourus si votre projet ne peut être finalisé à échéance et dans le budget prévu ? Comment comptez-vous pallier à ces risques ?***

Aucun risque n'est encouru dans la mesure où ces travaux entrent dans nos perspectives de recherche. Si le projet ne peut être finalisé à échéance ou que le financement est épuisé, nous pourrons ainsi continuer le projet tant que ce dernier ne sera pas achevé.

### ***6.12. Si vous demandez un soutien d'un an, accepteriez-vous de recevoir les crédits l'année prochaine plutôt que cette année ?***

----

**6.13. Si cette demande concerne la phase 1 d'un projet prévu sur 2 ans, pouvez-vous réaliser le projet en entier sur une année si vous obtenez les crédits en une seule fois ? Comment cela impacterait-il votre projet ?**

Non. Le projet ne pourrait pas être mené à terme car une phase incompressible de conception sera nécessaire en début de projet.

## 3. Architecture de l'outil

**6.14. Décrire l'architecture envisagée pour votre outil. Identifier les composants clés de l'application et décrire comment ils interagissent entre eux et avec les ressources matérielles.**

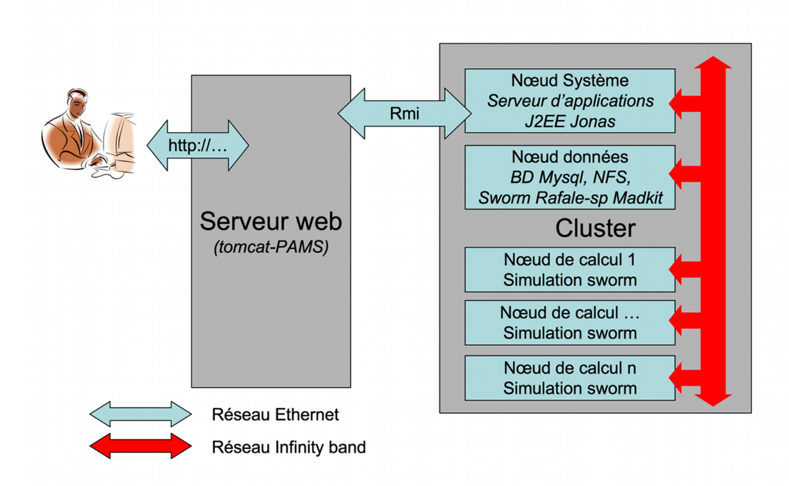
(*un schéma peut être appréciable*) schéma client -> accès

La première étape d'analyse du projet s'est conclue par la définition d'une structure logicielle simple et extensible, reposant sur des outils reconnus dans le domaine des systèmes répartis. L'architecture logicielle se décompose en trois éléments clés :

- **Le modèle Sworm** dans sa version distribué ainsi que Rafale-sp et Madkit pour son fonctionnement.
- **Une base de données** pour le stockage des caractéristiques des modèles (liste et format des paramètres d'entrée et des sorties).
- **Un serveur d'applications J2EE** permettant l'exécution des simulateurs sur des serveurs dédiés.
- **Une plate-forme web** pour la gestion de l'interface et permettant l'accès au modèle Sworm et permettant son utilisation (modification des paramètres, visualisation des résultats). Cette plate-forme se repose sur les environnements tomcat et PAMS

L'infrastructure de la plate-forme se compose :

- **d'un cluster** dans lequel un noeud est dédié à la gestion du système, un autre pour la gestion du disque et l'ensemble des autres pour l'exécution du simulateur
- **d'un serveur web** qui est à la fois connecté sur le réseau internet et sur le réseau privé du cluster.




**6.15. Lister les méthodes/référentiels d'analyses, de conception et de développement utilisés pour élaborer l'outil.**

Dans le cadre de ce projet, nous nous reposerons sur les méthodes traditionnelles de conception d'applications informatiques et sur des langages reconnus comme UML.

### ***6.16. Lister les langages de programmations, les bibliothèques parallèles et les outils de développement envisagés. Préciser le type de syntaxe qui sera utilisée pour la documentation du code.***

Les langages qui seront utilisés dans le cadre de ce projet seront Java, HTML, JavaScript, XML, … En outre, nous nous reposerons sur des plates-formes et bibliothèques issues des systèmes distribués telles que Jonas, Rafale-SP, MadKit, Tomcat, …

### ***6.17. Lister l'architecture matérielle (smp, cluster, grille de calcul…) et les logiciels requis pour faire fonctionner votre outil.***

L'architecture matérielle de l'outil repose sur des serveurs d'accès et un cluster de calcul. Le cluster de calcul n'étant accessible que depuis le réseau privé de l'université de Franche-Comté, il est nécessaire de compléter la solution par un serveur d'accès. De même, pour une plus grande liberté de configuration du serveur d'application et des bases de données, nous mettrons en place un serveur dédié. La virtualisation permettra de regrouper, sur un seul serveur physique, les composants qui ne seront pas exécutés sur le cluster. Du point de vue matériel, le cluster est composé de noeuds bi-processeurs interconnectés par un réseau de communication rapide, ce qui permet à la fois un mode de programmation en mémoire partagée et distribuée.

Les logiciels nécessaires à la mise en place de la solution sont soit destinés à la partie cluster, soit à la partie serveur d'application, base de donnée et serveur WEB. Pour la partie cluster, les principaux logiciels utilisés sont :

- Java pour le support de Madkit
- Les bibliothèques Madkit
- Le logiciel d'ordonnancement Sun Grid Engine

Pour la partie serveur d'application, base de donnée et serveur WEB, les logiciels utilisés sont :

- J2EE, serveur Tomcat
- La base de données MySql
- Serveur Web

### ***6.18. Comment ces choix influeront sur l'appropriation de votre outil par les utilisateurs cibles ?***

Ces choix n'influeront pas l'appropriation de l'outil par les utilisateurs dans la mesure où l'outil se présentera sous la forme d'un portail web. Toute l'infrastructure parallèle mise en place sera transparente pour l'utilisateur.

### ***6.19. Justifier le choix de ces technologies (conformité à des référentiels, robustesse, scalabilité, pérennité, communauté de développeur importante…) :***

La technologie J2EE et les serveurs d'applications comme Jonas et Jboss sont justement développés pour la construction d'environnements distribués robustes. Compte tenu des utilisations faites de ces outils dans le tissu industriel (service bancaire, jeux en ligne) leur pérénité et leur scalabilité n'est plus à démontrer.

Les outils utilisés dans le cadre de ce projet reposent sur une communauté de développeurs importante.

## ***4. Données en entrée et en sortie***

### ***6.20. Énumérer et décrire les données en entrée et en sortie de votre outil.***

Les données en entrée se limitent aux paramètres du simulateur qui sont saisis par le chercheur. Ils correspondent à des données physiques portant sur : (i) l'écosystème sol (granulométrie, quantité de matière organiques, etc) ; (ii) et la population de vers de terre qui évoluent dans ce sol. Ces derniers sont décrits par une taille, une espèce (*Millsonia anomala, Eudrilidae*) et des quantités journalières de sol ingéré.

Les données de sorties du simulateur correspondent : (i) à la quantité de matière orgranique présente dans le sol ; (ii) au volume occupé par les déjection des vers de terre, … Ses sorties se présentent sous la forme de données textuelles, de graphiques ou d'images en 2 ou 3 dimensions.

### ***6.21. Décrire la disponibilité (ou l'accessibilité), le format de stockage et d'organisation ainsi que la qualité des données utilisées en entrée. Quel est le coût et l'effort requis de l'utilisateur pour collecter, acheter, obtenir ou convertir ces données ? Dans quelles mesures le coût et l'effort requis limiteront-ils l'adoption de votre outil ?***

Les données en entrée se limitent aux paramètres du simulateur Sworm. Elles sont saisies par le chercheur utilisateur si bien que leur qualité est à la discrétion du thématicien.

### ***6.22. Les données seront-elles testées ou validées par l'outil en entrée ? Si oui, comment ?***

Des tests unitaires simples seront réalisés sur les données d'entrée (format de la donnée, borne de valeur). Cependant, la seule personne juge, capable de valider les données d'entrées, sera l'utilisateur, le thématicien en sciences du sol.

### ***6.23. Validerez-vous ou avez-vous déjà validé scientifiquement les données en sortie de votre outil ? Si oui, décrire comment cela se fera ou a été fait.***

Les données de sortie du simulateur Sworm ont déjà été validées scientifiquement et sont aujourd'hui reconnues par la communauté en sciences du sol.

### ***6.24. Décrire l'utilité immédiate des données en sortie de votre outil et les nécessaires conversions, post-traitements ou analyses ultérieures requis. Comment l'effort requis impactera-t-il l'adoption de votre outil par les utilisateurs cibles ?***

L'intérêt majeur du modèle Sworm est d'aider les écologistes du sol à comprendre comment un sol se structure sous l'effet des vers de terre. Cette structuration est particulièrement importante pour comprendre les flux d'eau, les activités microbiennes et les flux de nutriments (azote, phosphore) et de gaz (notamment gaz à effet de serre) associés. Il s'agit donc d'un outil important pour comprendre le fontionnement d'un sol (notamment toutes les fonctions dépendant de l'organisation physique d'un milieu-sol.

### ***6.25. Existent-ils des métadonnées ou y a-t-il production de métadonnées décrivant les lots de données en entrée ou sortie ? Si oui, comment sont-elles gérées et entreposées ? Sont-elles basées sur des standards ?***

Non

### ***6.26. La description ou le référencement des données est-il / sera-t-il basé sur un ou des référentiels ou thésaurus ? Si oui, lesquels ?***

Non

## 5. Interopérabilité

### 6.27. Quels sont les éventuels standards ou normes utilisées ?

Dans le cadre de ce projet, nous utilisons le standard J2EE pour le développement des composants logiciels distribués.

### 6.28. Votre outil est-il prévu pour être utilisé de manière interactive par les utilisateurs, par d'autres outils ou programmes (communication entre outils sur la base de requêtes ou autres) ou les deux ?

L'outil est prévu pour être utilisé de manière intéractive par les utilisateurs

### 6.29. Si votre outil pourra être utilisé dans les 2 cas, de manière interactive et de manière automatisée par d'autres applications, décrire les caractéristiques et fonctionnalités non accessibles pour chaque mode d'utilisation.

---

### 6.30. Si votre outil pourra communiquer de manière automatisée avec d'autres programmes, écrire brièvement 3 scénarios d'utilisation qui illustrent les détails de ces communications.

---

### 6.31. Si votre outil intégrera ou fera appel à des outils d'autres développeurs, décrire brièvement 3 scénarios d'utilisation

---

## 6. Rapports d'erreurs et d'avancement

### 6.32. De quelle manière votre outil montrera la progression du traitement aux utilisateurs ? Qu'est-ce qui sera signalé ?

A tout moment, l'utilisateur sera en mesure de se connecter à la plate-forme pour visualiser l'état de la/des simulations qu'il a lancées.

### 6.33. Comment votre outil notifiera-t-il à l'utilisateur l'apparition d'une erreur et quelles informations seront affichées dans le message d'erreur ?

Des messages d'erreurs seront affichés sur l'interface.

### 6.34. Avez-vous mis en place un processus de gestion des erreurs et de correction par l'équipe de développement et comment ?

Aucune gestion des erreurs n'a, pour l'instant, été mise en place à l'exception de celles utilisées par les développeurs.

## 7. Documentation

### 6.35. Quelles sont les différentes documentations prévues : nature et format de la (des) documentation(s) ? cible visée ? (spécifications fonctionnelles, spécifications techniques, docs/API développeurs…)

- Manuel d'utilisation du thématicien : comment utiliser l'interface web pour exécuter des simulateurs et coopérer ;
- Manuel d'utilisation du développeur : comment développer un simulateur compatible avec la plate-forme ;
- Documentation API du développeur ;
- Manuel d'installation de la plate-forme ;
- Articles scientifiques

### 6.36. Lister les sujets ou principaux chapitres qui apparaitront dans la/les documentation(s) de votre outil

**1. Objectifs du projet**
**2. Fonctionnalités proposées**
**3. Installation et configuration du système**
3.1. Installer le serveur web
3.2. Installer le serveur J2EE sur un cluster
3.3. Déployer le simulateur Sworm sur un cluster
3.4. Configurer le système et définir des politiques de distribution
**4. Expérimenter et analyser**
4.1. Réaliser une expérimentation en temps réel
4.2. Programmer des plans d'expériences
4.3. Récupérer et analyser des données de simulation
4.4. Contrôler l'avancement d'une simulation
**5. Concevoir et déployer un simulateur réparti**
5.1. Concevoir et développer un simulateur réparti avec RAFALE-SP et Madkit
5.2. Déployer un simulateur dans la plate-forme

## 8. Multilinguisme - traduction

### 6.37. Lister les langues parlées par vos utilisateurs cibles.

Les langues parlées par les utilisateurs sont le français et l'anglais.

### 6.38. Lister les langues dans lesquelles votre outil, votre documentation et tous les autres livrables seront traduits. Si vous ne traduisez par votre outil dans toutes les langues parlées par vos utilisateurs, comment cela affectera-t-il l'adoption de votre outil ?

L'interface de la plate-forme et sa documentation de la plate-forme seront dans un premier temps écrites en français. Ensuite, nous traduirons tous ces éléments dans d'autres langues.

### 6.39. Quelles méthodes ou technologies seront utilisées pour la traduction de votre outil, votre documentation et des autres livrables?

Aucune

## 9. Processus et équipe de développement

### 6.40. Avez-vous déjà géré des projets de développement logiciel précédemment ? Décrire brièvement votre(vos) expérience(s) passée(s).

La plupart des membres du projet sont des informaticiens et sont donc familiers aux techniques utilisées en génie logiciel.

### 6.41. Les développements seront-ils réalisés par des membres de votre équipe, par un prestataire sous contrat, ou autre ?

La création de la plate-forme nécessite une phase de développement importante. Cette dernière ne pouvant pas être supportée par des chercheurs qui ont déjà des activités annexes, elle sera confiée à des prestataires de service et/ou a des stagiaires rémunérés.

### 6.42. Si vous avez déjà sélectionné des développeurs, de votre équipe ou d'un prestataire, lister les, spécifier leurs rôles et décrire leurs compétences et leurs expériences passées. Attacher leurs CV si vous les avez.

Les prestataires de service seront sélectionnés dans un vivier d'étudiants en Master d'informatique inscrits à l'Institut de la Francophonie pour l'Informatique à Hanoi (Vietnam), l'Université Cheikh Anta Diop**,** l'Université de Franche Comté et à l'Université de Montpellier.

### 6.43. Si vous envisagez un prestataire de service, avez-vous déjà travaillé avec un prestataire auparavant ? Décrire comment vous vous assurerez qu'il développe ce que vous recherchez, dans les temps et avec le budget prévu.

Dans le cadre d'autres projets de nombreuses prestations de service ont été menées avec succès. Le contrôle de la bonne réalisation de ces prestations passe par des réunions hebdomadaires de projet visant à faire un état des lieux et à identifier le travail à fournir lors de la semaine suivante.

### 6.44. Impliquerez-vous vos utilisateurs cibles dans le processus de conception et d'implémentation de l'outil ? Si oui, décrire comment.

Un représentant des utilisateurs est complètement impliqué dans le projet. Il sera pour nous force de conseil.

### 6.45. Où sera hébergé le code source de votre outil durant son développement puis durant sa maintenance ?

Le code source de l'outil sera hébergé par le SIL de Montpellier. Un projet SVN existe déjà pour l'UMI Ummisco. Nous demanderons la création d'un sous projet dédié à EPIS.

### 6.46. L'outil sera-t-il placé dans une plateforme collaborative ou au sein d'une communauté de développement de projets open-source ? si oui, lesquels ?

Pour l'instant, nous n'envisageons pas le placement du simulateur sur une plate-forme collaborative comme source-forge. Cependant, cette option n'est pas à écarter si la communauté de développeurs autour du projet devient suffisamment grande.

## 10. Licence et distribution

### ***6.47. L'utilisation de l'outil sera-t-elle soumise à une licence pour les utilisateurs qui l'installeront sur leurs propres machines? S'agira-t-il d'une licence libre ? Le code source de l'outil sera-t-il protégé ou complètement ouvert ? (décrire l'éventuel coût, le type de licence et toutes autres éventuelles obligations)***

La plate-forme sera complètement Open-source.

### ***6.48. Existe-t-il des parties ou modules de votre outil qui sont protégés par des brevets ou des marques ?***

Non, aucune partie de l'outil ne sera protégée par des brevets.

### ***6.49. Décrire comment l'outil sera distribué ou rendu accessible aux utilisateurs (lister les sites web si nécessaire)***

Cette plate-forme sera (nous l'espérons) valorisée au travers d'articles en conférences et/ou revues d'informatique. Par ailleurs, les sources de la plate-forme seront disponibles et téléchargeables par tous via le site web du projet.

## 11. Installation

### ***6.50. La procédure d'installation sera-t-elle automatisée par un programme ou un script ou l'outil devra-t-il être installé « manuellement » ? (Préciser les OS et distribution)***

Pour l'instant, nous n'envisageons pas de procédure d'installation automatique de la plate-forme.

### ***6.51. Est-ce que le programme ou script d'installation détectera et signalera les logiciels requis manquants ?***

---

### ***6.52. Est-ce que le programme ou script d'installation permettra la désinstallation de l'outil ?***

---

### ***6.53. Si l'installation n'est pas pris en charge par un programme ou un script, existera-t-il une notice d'installation ?***

Oui, cette documentation fera partie des délivrables.

### ***6.54. De quelle manière la complexité de la procédure d'installation limitera l'adoption/l'utilisation de l'outil par les utilisateurs cibles ?***

Aucune dans la mesure, où les utilisateurs n'auront pas besoin d'installer un quelconque outil pour être en mesure d'utiliser la plate-forme. L'utilisateur accédera à l'outil via le web.

## ***12. Opération***

#### ***6.55. Les utilisateurs pourront-ils faire fonctionner l’outil sans votre aide? Si les utilisateurs doivent solliciter votre équipe ou des consultants externes ou suivre une formation, décrire les détails et les coûts***

Un des objectifs premiers de la plate-forme est de permettre à des non informaticiens d’assumer une activité de simulation à large échelle sans un concours constant des modélisateurs.

## ***13. Assurance qualité, maintenance et support***

#### ***6.56. Lister les techniques que votre équipe utilisera pour détecter les erreurs ou défauts.***

Afin d’avoir un meilleur suivi des erreurs, nous envisageons de définir une base de données des erreurs. Cette base stockera, de façon détaillée, l’ensemble des erreurs que les utilisateurs auront rencontrées ainsi que leur contexte. Cette base pourra, par exemple, stocker les exceptions JAVA qui surviennent durant une simulation ainsi que la valeur des paramètres de simulation.

#### ***6.57. Dans le cas où vous auriez un programme ‘beta’ en fin de développement, décrire comment il fonctionnera. Si des utilisateurs se sont déjà engagés pour l’utiliser, listez-les.***

La version béta de l'outil se présentera sous la forme d'un portail web. Pour l'utilisateur, aucune information sur le versionning de l'outil ne sera affiché explicitement. Les fonctionnalisés mises en place seront testées. Quand une nouvelle version de l'outil sera créé, le passage de l'une à l'autre sera transparente pour l'utilisateur.

Les utilisateurs qui sont engagés par l'utilisation de l'outil mis en place appartiennent à l'UMR Eco&Sol (Eric Blanchart, Jean-Luc Chotte, ...)

#### ***6.58. De quelle manière votre équipe fera-t-elle le suivi des erreurs dans ce projet ?***

La base de données des erreurs nous permettra de réexécuter les simulations posant problèmes et de mieux identifier les bugs avant de les corriger.

#### ***6.59. De quelle manière apporterez-vous un support à vos utilisateurs pendant la durée de ce projet, et après ?***

Le support aux utilisateurs sera effectué au travers d’outils comme des Forums, des Wiki, des tutoriels dont le contenu sera alimenté par une communauté qui, nous espérons, va se construire autour de l’outil. Nous pourrons aussi intervenir ponctuellement afin d’assurer un support technique fourni via la messagerie instantanée et les autres outils de collaboration de la plate-forme.

#### ***6.60. De quelle manière apporterez-vous un appui aux développeurs d’autres outils qui souhaiteraient utiliser et intégrer votre outil aux leurs ?***

Cet appui se fera sous la forme d’une documentation détaillée, d’une documentation API, d’un tutoriel montrant la démarche à suivre pour déployer un simulateur et l'exécuter sur le cluster.

## 8. Pertinence, résultats/livrables attendus et valorisation du projet

La demande peut-être être **accompagnée de tous documents utiles** :

*présentation du projet global ou descriptif du projet, rapport de phases préliminaires, étude de faisabilité (définition des besoins), dossier d'expression des besoins ou* ***cahier des charges****, devis détaillé…*

### *8.1. Résultats attendus (livrables) : (10 lignes maximum)*

Différents documents et outils informatiques seront livrés soit à l'issue soit durant le déroulement du projet. Ils se décomposent comme suit :

- *Outil* : Simulateur réparti Sworm
- *Outil :* Environnement réparti d'exécution de simulateurs sur un cluster
- *Outil* : Interface web pour le contrôle et la planification de simulations à exécuter sur un cluster
- *Outil* : la bibliothèque Rafale-sp
- *Outil* : L'environnement Madkit
- *Document* : Documentation complète de la plate-forme (utilisateur et développeur).
- *Document* : Documentation technique et scientifique portant sur la simulation distribuée
- *Document* : Résultats d'évaluation de la plate-forme sur des populations test.
- *Document* : Support de formation
- *Document* : Articles scientifiques réalisés dans le cadre du projet par les différents intervenants.

### *8.2. Pertinence du projet pour votre communauté scientifique*

Ce projet se place, en informatique, à l'intersection des systèmes distribués et de la simulation. Au regard de la littérature scientifique, les problématiques abordées dans chacun de ces domaines sont d'actualité et porteuses. La mise en place d'un environnement didactique, intuitif permettant la réalisation d'expérimentation distribuées et/ou parallélisées sur un cluster mettent en évidence des verrous technologiques qui sont loin d'être résolus. En ce sens, ce projet tente de répondre à un questionnement scientifique tout à fait intéressant et important pour la communauté scientifique en informatique.

Les résultats thématiques obtenus grâce à Sworm, l'intérêt notable que porte la communauté en sciences du sol vis à vis de cet outil montre le caractère inovant et l'importance de ce travail. Sa mise en ligne ne pourra qu'être profitable dans la mesure où Sworm pourra être appliqué à d'autres cas d'études que celui pour lequel le simulateur a été conçu (le sol de Lamto).

### *8.3. Pertinence du projet vis à vis des objectifs de SPIRALES / justification d'un financement DSI*

Ce projet répond à plusieurs objectifs de la DSI dont :

- **Répondre aux demandes thématiques en proposant des applications d'IS** : ce projet répond à la demande de biologistes souhaitant rendre accessible leur modèle sur internet.
- **Identifier les besoins et y répondre par des outils d'IS** : ce projet tend vers la construction d'une plate-forme pluridisciplinaire de simulation et d'analyse qui facilite l'accès à la simulation haute performance grâce à l'utilisation des technologies du web et des systèmes distribués.
- **Capitaliser les connaissances et les ressources** : L'outil mis en place favorisera la capitalisation des connaissances (résultats) et des ressources informatiques (cluster).

### *8.4. Retours sur investissement attendus (pour l'unité, l'institut…)*

Les retours sur investissement attendus vis à vis des laboratoires, des universités et des instituts partenaires sont importants car ce projet tend vers une mutualisation de la connaissance, des compétences et des moyens informatiques. A terme, la production scientifique (modèles, simulateurs, documentation, ...) et les moyens informatiques (clusters, plate-forme de simulation, outils collaboratifs pour la recherche) pourront être accessibles à partir d'un même point d'entrée (un portail web). Ainsi, toutes ces ressources scientifiques aujourd'hui éparses, méconnues et dupliquées pourront être rassemblées et partagées ce qui prévaut une amélioration de la productivité et une réduction des coûts.

### 8.5. Bénéfices du projet pour les pays du Sud

La mise en place et la publicité d'un environnement web permettant une réalisation massive de simulations et l'utilisation de ressources de calcul haute performance présente des avantages certains pour les partenaires des pays du Sud, pour plusieurs raisons :

- L'infrastructure nécessaire pour la mise en place d'un tel service est onéreuse (plusieurs centaines de milliers d'€uros). L'achat de ce type de matériels ne constitue pas une priorité pour les pays du Sud. Pourtant, l'utilisation de ces environnements devient de plus en plus habituel compte tenu des besoins grandissant en modélisation et en simulation de la complexité. L'outil que nous proposons permettra un accès à moindre coût à ces ressources informatiques, par nos partenaires du Sud.
- En mettant en commun, par le biais des NTIC, des ressources et des compétences limitées, les centres de recherche du Sud, pourront plus facilement atteindre une masse critique qui leur permettra de s'engager sur et même d'initier des programmes d'ambition internationale.
- La formation de chercheur en sciences du sol et de futurs chercheurs en informatique, respectivement par le biais, de séminaires et de stages de 6 mois.

### 8.6. Capitalisation, valorisation, transfert de savoirs-faire ou d'outils possibles ou prévus en matière d'IS

L'objectif de ce projet étant de concevoir et mettre en place une plate-forme pluridisciplinaire de simulation et d'analyse couplée à des outils de calcul haute performance, la capitalisation, la valorisation et le transfert de savoir-faire constitue l'essence même de ce projet. Ainsi ce projet tend vers une capitalisation à la fois de la connaissance (modèles, simulateurs, données de simulation, ...) et des moyens informatiques (le ou les clusters). Il s'inscrit donc dans une perspective de valorisation de tous ces moyens pour une meilleure utilisation. Par exemple, grâce au portail que nous souhaitons mettre en place, le mésocentre aujourd'hui sous-utilisé pourra demain être davantage sollicité par nos partenaires du Nord et du Sud.

Compte tenu de la capitalisation des connaissances intrinsèque à ce projet, il y a aussi transfert de savoirs-faire. Par ailleurs, il y aura aussi un transfert de savoir faire par le biais de la formation que nous souhaitons réaliser en fin de projet.

### 8.7. Valorisation possible ou prévue

Le projet sera valorisé de plusieurs manières :

- Participation à des conférences et publication en conférence ou/et en revue
- Site web
- Formations de chercheurs en sciences du sol
- Extension de la plate-forme à d'autres modèles
- Utilisation de l'outil par des chercheurs appartenant à d'autres unité

### 8.8. Observations particulières :